

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:19:57 ; Search time 46.2838 Seconds
(without alignments)
4811.606 Million cell updates/sec

Title: US-10-002-309B-2_COPY_24_886

Perfect score: 4597

Sequence: 1 ADNNSAIYFNTSQPINDLQ.....AWNGQYLDKSPKSRVRYK 863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4597	100%	898	082882	082882 escherichia
2	1436	31.2	310	092AL1	092al1 escherichia
3	686.5	14.9	1002	09AGX6	09agx6 vibrio chol
4	571	14.6	1335	09KN18	09kn18 vibrio chol
5	571	12.4	875	16 Q8D5P4	Q8d5p4 vibrio vuln
6	290	6.3	957	16 Q9KTH2	Q9kth2 vibrio chol
7	166.5	3.6	1217	16 Q97GW3	Q97gw3 clostridium
8	161	3.5	1038	2 Q45780	Q45780 bacteroides
9	159.5	3.5	2367	2 Q46034	Q46034 clostridium
10	158.5	3.4	2204	5 Q8IL70	Q8il70 plasmodium
11	156.3	3.4	1157	16 Q97FE2	Q97fe2 clostridium
12	154.5	3.4	2367	2 Q9F931	Q9f931 clostridium
13	149	3.2	1284	2 Q8RQ09	Q8rq09 bacillus gl
14	148.5	3.2	2468	16 Q912M3	Q912m3 pseudomonas
15	148.5	3.2	3692	2 Q8FRB3	Q8frr3 fusobacteri
16	148	3.2	1021	2 Q46085	Q46085 clostridium

17	148	3.2	2937	16 Q8D990	Q8d990 vibrio vuln
18	147.5	3.2	2306	16 Q8P107	Q8p107 leptospira
19	147	3.2	1386	16 Q92DL0	Q92dl0 listeria in
20	146.5	3.2	1377	16 Q98Q28	Q98qz8 mycoplasma
21	146	3.2	1093	16 Q8EV69	Q8ev69 mycoplasma
22	145.5	3.2	2189	5 Q8IKV6	Q8ikv6 plasmodium
23	143.5	3.1	1395	2 Q9AISO	Q9ais0 staphylococ
24	143	3.1	1284	16 Q8EV71	Q8ev71 mycoplasma
25	141.5	3.1	2057	2 Q9RE05	Q9re05 leuconostoc
26	140	3.0	1135	1 Q9P9D1	Q9p9d1 uncultured
27	140	3.0	1196	17 Q8TS48	Q8ts48 methanosarc
28	139	3.0	1110	17 Q8PXT0	Q8pxt0 methanosarc
29	139	3.0	1530	16 Q8Y479	Q8y479 listeria mo
30	138.5	3.0	1270	16 Q8XAN9	Q8xan9 escherichia
31	138.5	3.0	1575	2 Q9LCH3	Q9lch3 streptococc
32	138	3.0	958	17 Q8TLL4	Q8tll4 methanosarc
33	138	3.0	2894	17 Q58791	Q58791 methanococc
34	137	3.0	5559	16 Q8ZKG6	Q8zkg6 salmonella
35	136.5	3.0	1457	2 Q9F289	Q9f289 yersinia pe
36	136.5	3.0	3705	2 Q9F285	Q9f285 yersinia pe
37	136.5	3.0	3705	16 Q8ZHA1	Q8zha1 yersinia pe
38	136.5	3.0	3710	16 Q8CZU2	Q8czu2 yersinia pe
39	136	3.0	1677	16 Q8RGN7	Q8rgn7 fusobacteri
40	136	3.0	2167	16 Q92EK5	Q92ek5 listeria in
41	135.5	2.9	2529	16 Q25579	Q25579 helicobacte
42	135	2.9	1700	16 Q8XB95	Q8xb95 escherichia
43	135	2.9	1927	2 Q54875	Q54875 streptococc
44	135	2.9	9904	16 Q8NMQ6	Q8nmq6 staphylococ
45	134.5	2.9	4727	2 Q8GR92	Q8gr92 mycoplasma

ALIGNMENTS

RESULT 1

ID 082882 PRELIMINARY; PRT; 898 AA.
AC 082882; Q9ZGU1;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Plasmid pO157 DNA, complete sequence.
GN L7031.
OS Escherichia coli O157:H7.
OC Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RX MEDLINE=98290540; PubMed=9628576;
RC STRAIN=O157:H7;
RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA Yutsudo H.C., Kubota Y., Yamauchi Y., Iida T., Yamamoto K., Honda T.,
RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
RA Shinagawa H.;
RT *Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
outbreak*;
RL DNA Res. 5:1-9(1998).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RX MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT *The complete DNA sequence and analysis of the large virulence plasmid
of Escherichia coli O157:H7*;
RL Nucleic Acids Res. 26:4196-4204(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RA Bruner W.;

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; AB011549; BAA31757.3; -
DR EMBL; AF074613; AAC70099.1; -
DR EMBL; Y11831; CAA72517.1; -
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 898 AA; 9548 MW; 3CIAE23E3EAE1FAB CRC64;
Query Match 100.0%; Score 4597; DB 2; Length 898;
Best Local Similarity 100.0%; Pred. No. 3e-294;
Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ADNSAIYFNTSOPINDLOGSLAAEVKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLRVPV 60
Db 36 ADNSAIYFNTSOPINDLOGSLAAEVKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLRVPV 95
Qy 61 KADDKTPVQVEARDNNKILGTLTYPPSSLPDTIYHLDGVPEGGIDFTPHNGTKKIINT 120
Db 96 KADDKTPVQVEARDNNKILGTLTYPPSSLPDTIYHLDGVPEGGIDFTPHNGTKKIINT 155
Qy 121 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDPLEGKMWRFVSSAGYS 180
Db 156 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDPLEGKMWRFVSSAGYS 215
Qy 181 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENNRITYAQHINSAEIPAHWIYVPLNL 240
Db 216 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENNRITYAQHINSAEIPAHWIYVPLNL 275
Qy 241 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 275
Db 276 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 310
Qy 301 INVNYAPLHLKEVMLPTGELLTMDPGNGWHSQTMQRIGKELVSHGIDNANYGLNSTA 360
Db 335 INVNYAPLHLKEVMLPTGELLTMDPGNGWHSQTMQRIGKELVSHGIDNANYGLNSTA 395
Qy 361 GLGENSHYVVAQAAHNSRGNYANGIQVHGSGGGGIVTLDSITGNFSEHVGNYGLG 420
Db 396 GLGENSHYVVAQAAHNSRGNYANGIQVHGSGGGGIVTLDSITGNFSEHVGNYGLG 455
Qy 421 HYVDGFGSVHRSAAENNTGWDGDKRFIPNFYPSQTNESKSLNNOCEPFDGKFGF 480
Db 456 HYVDGFGSVHRSAAENNTGWDGDKRFIPNFYPSQTNESKSLNNOCEPFDGKFGF 515
Qy 481 DAMAGGSPFAANRFTWYTPNSAIIQRFENKAVFDSRSSTGFSKWNADTOEMEPYEH 540
Db 516 DAMAGGSPFAANRFTWYTPNSAIIQRFENKAVFDSRSSTGFSKWNADTOEMEPYEH 575
Qy 541 IDRAEQITASVNELSESKMAELMAEYAVVYVHWMNGNWTNIIPTASADNRGSLTINH 600
Db 576 IDRAEQITASVNELSESKMAELMAEYAVVYVHWMNGNWTNIIPTASADNRGSLTINH 635
Qy 601 EAGYNSYLFINGDEKVVYSGYKSFYSDGQFKERDVVDTRARKPEQGVPTTLVGY 660
Db 636 EAGYNSYLFINGDEKVVYSGYKSFYSDGQFKERDVVDTRARKPEQGVPTTLVGY 695
Qy 661 DPEGTLSYIYPAMYGAYFTYSDSQNDSDNCQLOVDTKEGQLRFLANHRANNVTYN 720
Db 696 DPEGTLSYIYPAMYGAYFTYSDSQNDSDNCQLOVDTKEGQLRFLANHRANNVTYN 755
Qy 721 KFHINVTESQPTQATLVCKNKKILDTKSLTPAPEGLTYVNGQALPAKENECCIVSVNSG 780
Db 756 KFHINVTESQPTQATLVCKNKKILDTKSLTPAPEGLTYVNGQALPAKENECCIVSVNSG 815
Qy 781 KRYCLPVGQSGYSLPDMIVGQEVYVDSGAKAKVLLSDWNLNSYNRIGEFVGNVPADMK 840
Db 816 KRYCLPVGQSGYSLPDMIVGQEVYVDSGAKAKVLLSDWNLNSYNRIGEFVGNVPADMK 875
Qy 841 KVAWNGQYLDSEKPSMRVYK 863
Db 876 KVAWNGQYLDSEKPSMRVYK 898
```

RESULT 2

```
Q9ZAL1 PRELIMINARY; PRT; 310 AA.
ID Q9ZAL1;
AC Q9ZAL1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 34.0 kDa protein (Fragment).
OS Escherichia coli.
OG Plasmid p0157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RA Brunder W.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
EMBL; Y11275; CAA72142.1; -
KW Hypothetical protein; Plasmid.
FT NON_TER 310
SQ SEQUENCE 310 AA; 34032 MW; 2AA72C166F315BA6 CRC64;
Query Match 31.28; Score 1436; DB 2; Length 310;
Best Local Similarity 100.0%; Pred. No. 1.2e-86;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ADNSAIYFNTSOPINDLOGSLAAEVKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLRVPV 60
Db 36 ADNSAIYFNTSOPINDLOGSLAAEVKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLRVPV 95
Qy 61 KADDKTPVQVEARDNNKILGTLTYPPSSLPDTIYHLDGVPEGGIDFTPHNGTKKIINT 120
Db 96 KADDKTPVQVEARDNNKILGTLTYPPSSLPDTIYHLDGVPEGGIDFTPHNGTKKIINT 155
Qy 121 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDPLEGKMWRFVSSAGYS 180
Db 156 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDPLEGKMWRFVSSAGYS 215
Qy 181 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENNRITYAQHINSAEIPAHWIYVPLNL 240
Db 216 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENNRITYAQHINSAEIPAHWIYVPLNL 275
Qy 241 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 275
Db 276 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 310
RESULT 3
Q9AGX6 PRELIMINARY; PRT; 1002 AA.
ID Q9AGX6;
AC Q9AGX6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ToxR-activated gene A protein.
CN TAGA.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=395;
RA MEDLINE=21117025; PubMed=11179381;
RA Karacalis D.K.R., Lan R., Kaper J.B., Reeves P.R.;
RT "Comparison of Vibrio cholerae Pathogenicity Islands in Sixth and
RT Seventh Pandemic Strains";
RL Infect. Immun. 69:1947-1952(2001).
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL; AF325733; AAK20748.1; -
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR SMART; SM00060; FN3; 1.
```

[illegible]

RESULT 4

```

Db      757 QMDQETORYKAVETDTPKPOQGVPPVVTLLGIYDPQENPSQIYPLVYSNGYVFELPQP 816
QY      558 KMAELMAEYAVVVMNGWNRNIYIPTASADNRGSILITINHEAGYNSY--LFINDEK 615
Db      817 EQEYQLE-----CW-----QAAGD-----LT-QAEIQYNQWOTLLIDGOQL 882
QY      616 VV-----SQYKKKSFVSDGQFWMKRDVVD--TREARKPEQFGVPVVTLLVGYDPEGTLS 667
Db      853 PICRFDYTNNGQSATEV--GSLNAQRNVCESRDMR-----WYNDYQIDSPVG 899
QY      668 SYIYPAMYGAIFYSDSQ-----NLSND----- 693
Db      900 QYELLSFGAGNYTTPNAEIGEVLCTLNKPHNNGSRDGGAGFVRNRCQEVGCKNNAE 959
QY      694 -----COLQVDTKEGOLRERLANHRANNTVMKRFHINPTE- 729
Db      960 GRVMSYAIRNSEVLSRTLASORCELVEHNRNGSTHALDGNRHKSTESNKFHNLSEK 1019
QY      730 SPTQATLVNCKKILDTKSLTPAPEGLTYTVNGQALPAKENECCIVSVNSGKRYCLPVGQ 789
Db      1020 GVPQVSLSCSD-ENGTSTLT-----RFTPD-QNPPLDKLKGPII-----IGQ 1060
QY      790 RSGYS 794
Db      1061 EYGYS 1065

RESULT 5
Q8D5P4
ID Q8D5P4 PRELIMINARY; PRT; 875 AA.
AC Q8D5P4;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN VV20864.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016811; AAC07787.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 875 AA; 97611 MW; ABEE604A79798C96 CRC64;

Query Match 12.4%; Score 571; DB 16; Length 875;
Best Local Similarity 27.3%; Pred. No. 6,4e-29;
Matches 182; Conservative 99; Mismatches 215; Indels 170; Gaps 27;

QY 236 PGLNLVI-----KQNLSGRLNDIKIGAPBELLHTIDIGMLTTPRDFPAKD-KEAHR 289
Db 3 PGLSLRLDNLGREGVLS--QGEIFGGAPELTIQNDGMGLIEPRNRTMIQNLPTLAA 60
QY 290 EYFQTIPIVRMIVNYPALHLKVEYMLPTGELLTDMDCNGGWSHSTNRQIRIGKELVSHGI 349
Db 61 DYFOKIPASKLVADYTPPAHPPIVTMPNGSVVYTKSASTGNGHSGDMREAIIGKAVYSTGI 120
QY 350 DNANTGLNSTAGLGEN-----SHPVVYVAQLAAHNSRGNYAN-----GIQVHGSGGGGI 398
Db 121 NNANVGILTSAGYSQYQYRRNH-----ITAHNVGYTTKKDLDLPQIVHSGSGGGGI 174
QY 399 VTLDTLGNESFHEVGHNYGLGHYVDGPKGSVHRSAAENNNSTWGDGKKRPIPNF----- 454
Db 175 VTEATTGNEWSHELGHNYGLGHM--PYMASIH-----DWESGWGDAPHQRFIGNLHWK 228
QY 455 ---YPSQTKEKSLNQCQEPF--DGHKFFGDAMAGGSP--FSANRFTMYTTPNSSAIIORF 509

```

```

Db      229 DYTQOQGD-----DIVPPFKDAFRFLVDAQNGGQEQYVGTISRTLEHHPAQSRKAQRM 282
QY      510 FENKAVFDSRSSTGFSKWNADTQEMEPYEHTIDRAEQITASVNEL-----SESKMA 560
Db      283 MNGFNLDSSSGYQWQAAQRYQTVETDTPKPOQTGVAVVNTLLGIYDPNENPSQI- 341
QY      561 ELMAEYAVVVMNMN-----GNWTRNIYIPTA-----SADNRGSILITINHEAGY 604
Db      342 -----YPLVTSNYGNVVELPQGVQAFQEGWQPVADLTPAELESDSWQTLRMDGQQRV 396
QY      605 NSYLF--INCDEKVVSGYKKKSFVSDGQFWMKRDVVD--TREARKPEQFGV--PVTLLVGY 660
Db      397 CKFTFOAANGDSAVFVGGVDQS-----TDRCSSGRDLQMHNSNMTSAQGDY 443
QY      661 DREGTILSSYIYPAH-----YGAIFYSDSQ----- 690
Db      444 E-----LLSKYGRGAVTYTTPPEVGEVTLCTLNKSGTDHGDGAGFVVGNNCEQISGVHKKHG 500
QY      691 -----DNDCOLQVDTKEG-----QLRERLANHRANNTVMKRFHINPTE 729
Db      501 TWRYALRGDEVLRPSYQTOGQCQLDVEFANGASERVVVLNASHSVNDS--NKFHVNLA 558
QY      730 S-QPTQATLVNCKKILDTKSLTPAPEGLTYTVNGQALPAKENECCIVSVNSGKRYCLPVG 786
Db      559 NGVPTQVRLSCSDRNGETELTRETPE-----QNPPIADLKGPII-----IG 599
QY      789 QRSYS 794
Db      600 QEYYS 605

RESULT 6
Q9KTH2
ID Q9KTH2 PRELIMINARY; PRT; 957 AA.
AC Q9KTH2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hemolysin-related protein.
GN VC0930.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004176; AAF94092.1; -.
DR TIGR; VC0930; -.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; 1.
KW Complete proteome.
SQ SEQUENCE 957 AA; 103989 MW; 4C6B40FBC755CDAD CRC64;

Query Match 6.3%; Score 290; DB 16; Length 957;
Best Local Similarity 53.8%; Pred. No. 2,4e-10;
Matches 57; Conservative 14; Mismatches 35; Indels 0; Gaps 0;

QY 756 LTYTVNGQALPAKENECCIVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSGAKAVL 815
Db 13 LLSLSNVQATTNAEAGCIIISRLNGEKYCLKVGERSGYSLPSWIYAHVPDVQAPSGVSM 72

```

```

816 LSDWDNLNSYRNGIFGVGNVPADMKVKAWKNCQYLDFSKPSMRVV 861
||||||| : | ||||| ||||| : 
73 LSDWDNLNSYRLAVFDRTYTGTNEDLNKKAYNGA YLDFS KPSMRYL 118

RESULT 7
ID Q97GW3 PRELIMINARY; PRT; 1217 AA.
AC O97GWS;
DT DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE DE Alpha-glucosidase fused to unknown alpha-amylase C-terminal domain.
DN GENE
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxId=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zentgraf T., Glaser S.M., Hitt D.L., Wolf Y.I., Soucaille P., Tatunov R.N., Sabathe F., Doucet-Stamm L., Bennett G.L., Koonin E.V., Smith D.R.; "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."; J. Bacteriol. 183:4823-4838(2001).
RL EMBL; AE007725; AAK80209.1; -.
DR InterPro; IPR005085; CBM_25.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF03423; CBM_25; 1.
DR Pfam; PF01055; Glyco.hydro.31; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW Complete proteome.
SQ SEQUENCE 1217 AA; 134530 MW; 26F9D66B02688A30 CRC64;
```

Query Match	3.6%;	Score 166.5;	DB 16;	Length 1217;
Best Local Similarity	18.6%;	Pred. No. 0.049;		
Matches	179;	Conservative 121;	Mismatches 319;	Indels 343; Gaps 48;
Qy	56	LVRPVA-----DDKTPVQVEARDONNKLGLTLLTPPSSLP-----DTIYHLDGVPEGGI	106	
Db	29	IIRPVSAKAPNMATISKIKTVKENARV-----SNLSAKLNGDPLQIVNGLDETDI	79	
Qy	107	-----DFTPHNGTKKIINTVAEYVKNKLSDAGSSIHSHLNTNALV-----EHTA	150	
Db	80	KICEPOVLKVDYKP-SGQSSSDTLIVDPNPKIWN-TGNIISDLNDSMPVITTTQKMTIKIS	137	
Qy	151	NGRWVRDIYLPQGDPLECKMY---RFVSSAGYSSTVFYGDGRKVTLSVGTNTLLFFKVVNGQW	207	
Db	138	KSDLTMSVY-----DSTGQIVKQOOSIASKSVSFTHNSGDR-----FYGINCYN	181	
Qy	208	FR-----SGELEN-NRITYAQHIHWSAELPAHIVPGLNLIIVIQGNLSGRND-----	253	
Db	182	FKEDSNKGMRLNGTESYVYQGHGCSFPVWSDGYGLLVSDGGSFTIGTSLQYSGIS	241	
Qy	254	-----IKICAPGELLHTTDI-----GMLTTP-----RDRFDFAKOKEAHRE	290	
Db	242	KTDTDYILMLSPKEVISESDVSKAPFPKATGTFTNTQWNNNSLSCGTGNDEDKLKS	301	
Qy	291	YFQT-----IPVSRMIV-----NNYAPLHLKEVMLP---TGEILTMDPPGNGGWHHS	333	
Db	302	VLNTRYRSQIPIDNFCFLDFENKKMGQODNYGEFKWNTDNFPDAQNGQLKAYMD-SKGLKMT	360	
Qy	334	GTMORRI-----KELYSHG-----IDNANTGLNSTAGLGENSHPYV	370	
Db	361	GIMKPRIILADSEQARYVTSKGWMLPGDSAAQSDYCSGKKMHEWNVFAISDV-----	409	
Qy	371	VAQLAAHNSRNYANGIOVHGSGGGGGVTTLDTSLGNFESHEVGHNYGLGHY-----	422	

Db	410	-----RKWNWNNIO-----DAFDKGIVGTFWDECDENVNTGNGMMNMAERA 450
Qy	423	-YDGFSGVHRSAAENNNSTWG-----WDGDKRFRIPNFPYSPOTNEK-----462
Db	451	LYDG-----QRAYKNQRVSLNRNYAGAQRYCYGMWSGD-----ISTGFDSDMANQRMR 500
Qy	463	-SCLNNQOEFPDGHKCFD--ANAGSP-----FSAANRFTMYTPNSSLAIQHFENKAV 515
Db	501	LSAVN-----LGEAKWGMGTGGFNDGDPPTENYARWMEESAFTP-----IFRVHGODNRV 550
Qy	516	-----FDSRSSTGFSKWNADTOEMEPYEHTIDRAEQITA-----SVNELSESK 558
Db	551	RYPWAFGSTAEAAAKKANQLRYTLPIIYYSVDRSASQSGGLYVSLMMEYPNDSNAANDK 610
Qy	559	MAELMAEYAVVYKHMNGNWTNRTNYP-----TASADNRGSGILITINHEAGVNSY-----607
Db	611	EAWMEGDYMLVSPVYNQGTQSKSYLPGNNIDYTTGREYTGQ--QTINAVDSTNNMSDIP 669
Qy	608	LFING-----DEKVVSGYKKFSVSD-----GQFWKKER 635
Db	670	LFIKSGAIIPTODFENYGEKKIITDYYVDAPFSDKATTFDYDDDGTSYDENGSTFDDK 729
Qy	636	DYVDTREARKPEQFGVPVTTLVGYDPEGLSSYI-----YP 672
Db	730	MTLQTSSTDSKSVQFNIDKNT--GSTPDP--LKDYIVKMHYKNGCAVTVANGOALTOYSSYD 785
Qy	673	AMYGAYGFTYSDSDSONLSNDQCLOVDTKEGQLRFLRANRANTYVANKHINVPTESQP 732
Db	786	ALKSASGEGYASGT-----DTYGNVYVTKYSSGDAKN-----INVSCNPLP 826
Qy	733	TOATLVCNNK-----ILDTKSLTPAPEGLT--YTVNGOALPAKENEGCIVSVNSGKRYCL 785
Db	827	VITIAAANPKFGTYGTPQTVSLTASKSDATYYTL-----DGTTFVNSTK-YIA 875
Qy	786	PV 787
Db	876	PI 877
RESULT 8		
ID	Q45780	
IT	Q45780	PRELIMINARY; PRT; 1038 AA.
DT	Q45780;	
DT	01-NOV-1996 (TrEMBLrel. 01, Created)	
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Outer membrane protein.	
GN	SUSC.	
OS	Bacteroides thetaiotaomicron.	
OC	Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;	
OC	Bacteroidaceae; Bacteroides.	
OX	NCBI_TaxID=818;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BT55482;	
RC	MEDLINE=96146534; PubMed=8550519;	
RA	Reeves A.R., D'Elia J.N., Frias J., Salyers A.A.;	
RT	"A Bacteroides thetaiotaomicron outer membrane protein that is	
RT	essential for utilization of maltooligosaccharides and starch.";	
RL	J. Bacteriol. 178:823-830(1996).	
CC	- - SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).	
DR	EMBL; L49338; AAA95938.1;	
DR	InterPro; IPR001395; Aldo/ket.red.	
DR	InterPro; IPR000531; TonB_boxC.	
DR	Pfam; PF00593; TonB_dep_Rec; 1.	
DR	PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.	
DR	Outer membrane; Outer membrane; Receptor; TonB box.	
SW	SEQUENCE 1038 AA; 115886 MW; DD4C6CE38A2DC5E1 CRC64;	

Query Match 3.5%; Score 161; DB 2; Length 1038;
Best Local Similarity 18.1%; Pred. No. 0.088;
Matches 180; Conservative 120; Mismatches 328; Indels 36

```
QY 16 NDQGSIAAEVKEFAQSILPAHPREGDSOPHLASRLKSLILVR-----PVKADDKTP---V 68
DB 131 NDMTGSVMA-----IKDELKSG-----IITNAQDMLSGKIAGVSVISNDGTPOGGA 177
QY 69 QVEARDNNKILGTLTLYPSSLPDIYHLDGVPGEIGDTPHNGTKKIINTVAENK--- 126
DB 178 QIRIRGSSSL-----NASNDPLIVIDGL---AID---NEGIKGMANGLSMWNPAD 221
QY 127 -----LSDASGSIHSLTNALVEIHTANGWRVRIYLPQGLGKMWREVSSAGYS 180
DB 222 IETLVTKDASATAIYGSRASNGVIITTKKG-----NGQAPSVYTINGSVSFS 270
QY 181 STVYFGDKVTLSVGNLTLLFKYNGQWFRSGELENRRITYAQHIWSAELPA-----HW 233
DB 271 KT-----QKRYDVLSDGEYRA-----YANQLMGDKLPADLGTANTDW 307
QY 234 -----IVPGLNLVTKQG--NLSGRNLNDIKIGAPGELLHLTHIDIGMLTPDRDFPAK 283
DB 308 QDQIFRTAVSTDHHSVINGGFKNLPRYS-----LGYTDDNGIVKTSNFR-RFTA 356
QY 284 DKEAHREYFQ---TIPYSRMIV---NNYAPLHLKE---VMLPTGELLTMDP---GNGG 330
DB 357 SVNLAPSFEDHLKFNINAKFMKNGKNRYADSRKYWRALAIIDTPRPYSNEDDPQFTGGY 416
QY 331 WHSGTMRQRIKELVSHGIDNANTGLNSTAGLGE-----NSHPYVVAQLAAHNSRG 381
DB 417 WQN-----INSTGFSNPDMKYTSNPNSSPONPLAALKLNKDKG 454
QY 382 NYANGIQVHGGSGGGIVTLDSTLGNFSEHVEHNYGL-----GHYVDGFKG 428
DB 455 N-----SNDYFGVNDVYKFEHFLPDLRLHASIGGEYAEQTOT 491
QY 429 SVHRSANNSTMGWGD---KKRFTPNFVPSOTNEKSLNOCQEPFDGKHGFGDAMA 484
DB 492 TIVSPYSEGNYYGNGDVTQYKYNLSYIVQYI--KSLGAND-----FLIMV 538
QY 485 GGSFSAANRFTMYTPNSSAIQIRFFENKAVDSRSSTGFSK-----WNAOTQEMPEYHTI 541
DB 539 GGE-----QHFRN-----GPEEGQGWDSYTOE-----PHDA 566
QY 542 DRAGQITASVNELSESKMA-----ELMAEYAVVVMHNGWNRNIIYPTASADNRGSILT 597
DB 567 KLREQATAYATRLTVSYFGRNLNLLARYLFTFTMRWDGS-----SRFSKDNRW--- 615
QY 598 INHEAGYNSYLFINGDEKVVSGYKFSVSDGQFWKERDVVDTRARKPEQFGVPVTLV 657
DB 616 -----GTPPSLALGKH---IKENFLKDVNLVSLDKLR---LHGWIQQQ 654
QY 658 GYDPEGTLSYIYPAMYGAYGF--TYSDDS--QNLSDNDCQLQVDTKEGQLRFLANHR 713
DB 655 NIGDDFAYLPLYVYVNNYEAQYPPGDTYYSTSRPKAFENLKWETTTWNAGLDFGFLNGR 714
QY 714 ANNTVMKHEHNVTESQPTQATLVCNKKILDTSKPAPESGLTY-----TYNGQ 763
DB 715 ITGIDGYF-----RKTCTVTLRSPNNILOCPDDEYRFTGKLYRGSINAK 763
QY 764 ALPAKE-----NEGIVSVNSG--KRYCLPVG----- 788
DB 764 PIVTKDFTWDLSYNITWNHNEITKLTDGSDSYVEAGDKLSRGNNTKVQAHKVGYYAANS 823
QY 789 -----QRSQSYSLPDWITGVQVYVDSGAKARVLLSDNLSYNRIGVEFGVNVN 835
DB 824 FVYSRGNNTKVQAHKVGYYAANSFYVYQVYDENGKPI-----ENMFVDRNGN--GTID 874
QY 836 PADMKKVKAWNG-----QYLDTSKPKRSMR 859
DB 875 SGDKYIYKKPAGDVLMLGLTSKMQYKKNFDFSESLR 908
```

RESULT 9

Q46034

ID Q46034

PRELIMINARY:

PRT: 2367 AA.

```
AC Q46034;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Toxin B.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-isolate 1470;
RX MEDLINE=96079281; PubMed=7494480;
RA von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,
RA Sartingen S.;
RT *closing in on the toxic domain through analysis of a variant
RT Clostridium difficile cytotoxin B.*;
RL Mol. Microbiol. 17:313-321(1995).
DR EMBL; Z23277; CAAB0815.1;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW_binding_1; 18.
DR PROSITE; PS01118; SUI1_1; 1.
SQ SEQUENCE 2367 AA: 269186 MW: EF9823DAE70427F3 CRC64;

Query Match 3.5%; Score 159.5; DB 2; Length 2367;
Best Local Similarity 18.4%; Pred. No. 0.4;
Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;

QY 4 NSAIYNTSOPINDLOGSLAAEVKEFAQSILPAHPREGDSOPHLASRLKSLILVRPVKAD 63
DB 1454 NSELQKNIPYFVDSSEK-----ENGFINGSTKEG-----LFVSELPDVLISKVYMD 1501
QY 64 DKTP-----VOVEARDNNKILG-----TLTLYPSSLPDIYHLDGV 101
DB 1502 DSPSPGYYSNNLKDVKYITKDNVNLITGYLLKDDIKISLSLTLODEKIKLSNVHLD-- 1559
QY 102 PEGGIDFTPHNGTKKIINTVAEYVKNLSLDASGSIHSLTNALVEIHTANGWRVRIYLP 161
DB 1560 -ESG-----VAEILKPMNRKGS---TNTSDSLMSFLESNN---IKSIFV- 1596
QY 162 QGPDLEGKMYRFVSSAGY--SSTVYFG-----DRKVTLSVGN 197
DB 1597 ---NFLQSNKFTILDANFIISGTSIGQFEFICDENNNIQPYFIKNTLTNTLYVGNR 1653
QY 198 L-LFKYVNGQWFRSGELENRRITYAO-HWSAELPAHWIYVGLNLVKNQNLGS----- 249
DB 1654 QNMIVPEYDLDSDGDISSTVINFPSQKLYLGIDSCYKVVISNIIYTDENIITPYETNN 1713
QY 250 -----RLNDIKI-----GAPGELLHTIDIGMLTPDRDF-DFAK 283
DB 1714 TYPEVIVLDANYINEKINVINVDLSIRYVNSDNDNFILMSTSEENKVSQVKIRFVNVFK 1773
QY 284 DKE-AHREYF-----QTIPYSRMIVVNNYAPLHLKEVNL----- 315
DB 1774 DKTLANLSFNSDKQDVPVSEIIL-SFTPSYEDGLGIDGLVSLYNEKFIYNNFGMM 1832
QY 316 -----PTGELLTD-----MDPGNGGWHSGTMRQRIKELVSHGI 349
DB 1833 VSGLIYINDSLYFKPPVNNLITGFTVVGDKKYFNPINGGAAS-----IGETI----I 1862
QY 350 DNANYGLNSTAGLGENSHPYVVAQLAAHNSRGNY-----ANGIQVHGGSGGGIYIDST 404
DB 1883 DDKNYFNQSG-----VLQTVGFSTEDGFKYFAPAN-----TLDEN 1918
QY 405 LGNE---FSHEVGHNTYGLHYVDGFKGSHRSANNSTNGW---DGDKKRFPINFPYQ 458
DB 1919 LEGEAIIDFTCKLIIDENIYFEDNYRGAVE-----WKELDGEHMFYSPE----- 1962
QY 459 TNEKSLNQCQEPFDGKHGKFGFDAMAGGSPFSAANRF--TMYTPNSSAIQIRFF-----EN 512
DB 1963 -----TGKAFKGLNQIGDDKYFNDSGVNMKGFGVSINDN 1996
```



```
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE-21359325; PubMed-11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabatche F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT *Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum;
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007786; AM80833.1;
DR InterPro; IPR005085; CBM_25;
DR InterPro; IPR000322; Glyco_hydro_31;
DR InterPro; IPR000531; TonB_boxC;
DR Pfam; PF03423; CBM_25; 1;
DR Pfam; PF01055; Glyco_hydro_31; 1;
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1;
KW Hydrolase; Complete proteome.
SQ SEQUENCE 1157 AA; 129411 MW; 1DB17280833F7578 CRC64;

Query Match 3.4%; Score 156.5; DB 16; Length 1157;
Best Local Similarity 19.3%; Pred. No. 0.21;
Matches 198; Conservative 120; Mismatches 335; Indels 371; Gaps 56;

QY 60 VKADDTTPVQVEARDNNK-----ILGTLTLYPPSSLPDTIYHLDGVPEGGIDTPNGTK 115
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 35 VKADTNKVNKNSKSSQKFKHAKNGTLTKIKKGDETIIRICEPQVKVDYKP-NGKS 93
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 116 KIINTVAEVNKLSDAGSSHTSHLTNNAIVEHTANGRWYRDI-----YLPQGPDL 166
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 94 SKDTLVDPNK-KWSTGNIVSSDIKSDPWV---ITTKKWLKLNKEDLSLVY-----DL 144
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 167 EGMKRVFSSAGYSSVVF---YGRKVTLVSGNTLLFKYVNGQWFR-----SGELEN-NRI 218
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 145 QGKLLKQDTSKASTASTHNSGR-----FYGINGYFQEDSSKGLMRNGTES 193
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 219 TYAQHIWSAELPAHWIYVPLNLYIKQNLGRLND-----IKIGAPGEL 262
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 194 VYAGYQHGCGSPFVMSWDGVLVDSGGSFITGDTSLKYDGISKDTDTYVVMVGNPKEI 253
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 263 LLHTIDIG-----MLTTPDRF-----DFAKDKAEHREY-FOTIPVSRMI 301
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 254 LSESDVSGRAPPPKWANGFTNTQGWDNLSLSTGNDKAKLSVINTYRSKQLPIDNFC 313
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 302 V-----NNYAPLHLKEVLP---TGELLTMDPGNGWHSQVNRQRI-----GRE 343
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 314 LDFDKKKGODNYGEFKWNTDNFPDSQGLKAYMD-SKGLKMTGIMKPRILADSKQRY 372
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 344 LVSHG-----IDNANYGLNSTAGLGENSHPYVVAALAHNSRGNFYANG 386
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 373 VTSKGWNLPGDSEASDYCSKMMENVPAL-----PQV-----KKMMNN 412
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 387 IQVHGSGGGGIVTLDTSLGNEFSHEVGHNYGLGHY-----VDGFKGSVHRSANN 437
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 413 IQ---GAFDKGLVGF-----WNDECDENVNFCGNMNERAIYDS-----QRRHKN 456
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 438 NSTWG-----WDGDKRFIPNFPVPSOTNEK-----SCLNQCQEPFPGHK 477
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 457 QRVNLSARNYAGAQRYSGWMSGD-----ISTGDSMANQRERMLSAVN-----LGEAK 506
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 478 FGFD-AMAGSP-----FSAANRRFTWTPNSSAIIQRF--ENKA-----VFDRSSTGFS 525
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 507 WMDTGFNGGDPPTPENYARWESAFTP-----IPRVHGDNKRVPWPWFGSTAETAK 561
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 526 KWNADTOEMPEYHTIDRAEOITA-----SVNELSEKMAELMAEYAVVKVHM 573
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 562 KAMQLRYTLIPYISYDRSASQGLVRSILMMEYPNDSNAANDKEAMFGDYMLVSPV 621
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 574 WNGWNRNIYP-----TASANDRSILTNHEADGNSY-----LPIG----- 612
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 622 QEGOTSKSIYLPENGWIDYTTGREYTGQ-QTINAVDSTNWSIDPLFKSGAIPTQDFE 680
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 613 ---DEKVVYSGQYKKSF-----VSDGQFWKERDVTDFEARKPQFG 650
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 12
Q9F931
ID Q9F931 PRELIMINARY; PRT: 2367 AA.
AC Q9F931;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Cytotoxin B.
GN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5340;
RX MEDLINE=20448897; PubMed=10992443;
RA Sambol S.P., Merrigan M.M., Lyerly D., Gerding D.N., Johnson S.;
RT *Toxin gene analysis of a variant strain of clostridium difficile that
RT causes human clinical disease.;
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL; AF217292; AAG18011.1;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR001950; TIF_SUT1.
DR Pfam; PF01473; CW_binding_1; 18.
DR PROSITE; PS01118; SUT1_1; 1.
DR PROSITE; PS01118; SUT1_1; 1.
SQ SEQUENCE 2367 AA; 269337 MW; D5EE715E5B041E2F CRC64;

Query Match 3.4%; Score 154.5; DB 2; Length 2367;
Best Local Similarity 18.4%; Pred. No. 0.86;
Matches 185; Conservative 148; Mismatches 322; Indels 353; Gaps 49;

QY 4 NSAIYNTSQPTINDLOGSLAAEVKFAQSOILPAHPKSGSQPHLTSLRKSLLLVRPRAD 63
DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1454 NSELQKNIPSYFVDSRGK-----ENGFNGSTKEG-----LFSVSELPVVLISKVYMD 1501
DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 64 DKTP-----VQVEARDNNKILG-----TLTLYPPSSLPDTIYHLDGV 101
DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1502 DSKPFGYSSNNLKDVKITKDNVNILTYGKDDIKISLTLQDEKTIKNSVHLD--- 1559
DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 102 PEGGIDTPHNGTKRIINTVAEVNKLSDAGSSHTSHLTNNAIVEHTANGRWYRDIYLP 161
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1560 -ESG-----VAEILKPMNRKS--TNTSDLSMFLSNM---IKSIFV- 1596
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 162 QGPDLEGKMWRFVSSAGY--SSTVFG-----DRKVTLSVGMT 197
DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1597 ---NFLQSNKIFLDANFIISGTTISGTFEFCIDENNNTQPIFKFNTLETNTLYVGNR 1653
DB ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 198 L-LFKYVNGOWFRSGLENNRITYAQ-HIWSAELPAHWIYVPLNLYIKQNLGSG----- 249
DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```


1654 QNMIVPNYDLDDSGDISSTVINFSQKYLIGIDSCVKNVWISPNIIYTDENITPVYETNN 1713
QY 250
DB 1714 TYPEVIVLDANYINEKINYNINDLSIRYVNSDGNDFILMSTSEENKVSQKIREVNFVK 1773
QY 284 DKE-AHREYF-----OTIPVSRMIVNNTYAPLHLKVMVL----- 315
DB 1774 DKTLANKLSFNSQDQVPVPSLIL-SFTPSYEDGLGIDGLYSLYNEKFIYNNFGMM 1832
QY 316
DB 1833 VSGLIYINDSLYFKPPVNNLTIGFTVGDKKYFNPINGGAAS-----IGETI-----I 1882
QY 350 DNANVGLNSTAGLGENSHPYVVAQLAUNSRNY-----ANGIQVHGGSGGGIYTLDS 404
DB 1883 DKNYFNGSG-----VLQGVFSTEDGKYFAPAN-----TLDEN 1918
QY 405 LGNE-----FSHEVGHNYGLGHYVDGPKGSVHRSAENNSTWCW---DGDKKRFPNFPSPQ 458
DB 1919 LEGEALDFTGKLIIDENIYYPEDNYRGAVE-----WKELDGEHMFESPE----- 1962
QY 459 TNEKSLNNOQOPEPFGHKEFGDAMAGSPFSAANRF---TMYTPNSSAIQIRFF-----EN 512
DB 1963
QY 513 KAVFDSRS--STGFSK-----WNADQEMEPYEHTIDRAEQITASVNSLSESKMAELMA 564
DB 1997 KHYFDDGSGVKNVGYTIDGKHFFYFAENGEMQIGVFNTEDEGKYFAHHNEDLNGEEGEIS 2056
QY 565 EYAVYVHMNGWNTENIYPTASADNRGSILTIHNEAGNSYLFINGDEKVVSGOYKKS 624
DB 2057 YSGIL-----NFNNKIY---FDDSFYAVVGWKLDEGSKYF---DEDTAEAYIGLS 2103
QY 625 FVSDGQFWERDVVDTRARKPEQFG-VPVTLVGYDPEGTLSSLIYIPAMYGAVGFTYS 683
DB 2104 LINDGQYFNDGDI-----MOVGFVINDKVFYFSDSGIIES-----GVQNI 2145
QY 684 DQSNLSDNDQLOV---DTKEGQLFRLANHRANNTVMNKFHINVPTESQPTQATLVCN 740
DB 2146 DNNYFIDNGIYQIGVFTSDGKYFAPA-----NIVDNIY----- 2183
QY 741 NKILDTKSLTAPPE-----GLTYVNGQALPAKENEGCIVSVNSGKRYCLPVGQRS--GY 793
DB 2184 GOAVEYSGLVVGVGDYVYFGETYTIETGIWIDMENE-----SDKYFDPETKCKACKGI 2236
QY 794 SLDPNIVGQEVYVDSGAKAKVLLSDNDLSY--NRIGEF-VGVNVPAD 838
DB 2237 NLIDDI---KYFDEKGIHRTGLISFENNYYFNENGEMQPGYINIED 2281

RESULT 13
Q8R0U9
ID Q8R0U9 PRELIMINARY; PRT; 1284 AA.
AC Q8R0U9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 6-glucosyltransferase.
GN CTSZ.
OS Bacillus globisporus.
OC Bacteria; Firmicutes; Bacillales; Planococcaceae; Sporosarcina.
OX NCBI_TaxID=1459;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cil;
RA Maruta K.;
RT "Cloning and sequencing of the genes encoding cyclic tetrasaccharide-
synthesizing enzymes from Bacillus globisporus Cil.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB073929; EMBL: AB084041;
DR InterPro: IPR005084; CBM_6.
DR InterPro: IPR003322; Glyco_hydro_31.

DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF03422; CBM_6; 2.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Transferrase.
SQ SEQUENCE 1284 AA; 139128 MW; 2167B0DEB4F2E9C CRC64;
Query Match 3.2%; Score 149; DB 2; Length 1284;
Best Local Similarity 18.8%; Pred. No. 0.76;
Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;
QY 82 TLTLYPPSSLPDITYHLIDGVPEG--GIDFTPHNTKIIINTVAEVNKLSDASGSSIIHS- 138
DB 54 TLTVDNGAEPSSDOLLIVQAVQNGILKVDYRNSITPSAKTIPMLDPNKTWSAVGATINTTA 113
QY 139 -----LNNALVEI-----HTANGRWYDIYLPQDPLECKMYRFVSSAGYSTVF 184
DB 114 NPMITTSSMCKIEITKTPVMTVKADGTTL--FWEPSSGGGVFSDGVRFLHATGDN---M 168
QY 185 YGDRKVTLSVGNLTLLFKYVNGQWFRSGELENRRITYAOHWSAELPAHVIPLNLVIKQ 244
DB 169 YGIRSF-----NAFDSGGDLRNSNHAH----- 193
QY 245 GNLSRLNDIKIGAPCELLHTIDIGMLTTPRDFPAKREAHREYFQ---TIPVSRMIV 302
DB 194 ---AGEQGD---SGGLIWSTAGYGLLVSDGGYPTDSTTGOMEFFYGGTPEGRRYA 246
QY 303 NNYAPLHL-----KEYMLPTGELLTMDPGNGMHSHTMR-----Q 338
DB 247 KQVEYYITMLGTPREIMTDVGE--ITGKPPMLPKSLGPFNFEMWDTNQTEFTNNVDITYAK 305
QY 339 RIGKELVSHGIDNANYGLNSTAGLGENSHPYVVAQLAAHNSRGNANGIOVHG----- 391
DB 306 NIPIDAYAFDYDKKYGKGETNYGEPAWNTNPPSASTTSLKSTMD-AKGIKMIGITKPRIV 364
QY 392 -GSGGGGIVT--LDSTLNEFSHEVGHNYGLHYVDGPKGSVHRSAE---NNNSTWGHG 445
DB 365 TKDASANTVTOGTDTATNGGYF--YPCHN-----EYQDYFIPVTYRSIDPYNANERAWFNH 418
QY 446 DK---KRFPNFPSPQTNKS-----CLANCOQEPDFGHKFGFDAMAGSP--- 488
DB 419 STDALNKGIVGHNDDETCKVSSGGALYWFENFTTGHMSQTYEGOR-----AYTSAQRVW 474
QY 489 -----FSAANRF---TWYTP-----NSSAIQRFENKAVFDSRSSTGFSKN 528
DB 475 QTARTFYPCAQRATYATLWSGDIGIQNKGERINWAAHQ---EQRAVMLSSVYNGQVKG 531
QY 529 ADT-----QEMEPY-----EHTIDRA 544
DB 532 MDTGGFNQOOGTTNNPNPDLYARMHQFSALTPTVFRVGHNNHQRQRPWYFGSTAEEASKEA 591
QY 545 EQI-----TASVNSELSKMAELMAEY---AVVK---VHMVN-GNV----- 578
DB 592 IQLRYSLLIPYMYAYERSAYENGVLVRPLMQAYPTDAVKNTYDAMFGDWLLAAPPVVK 651
QY 579 -----TRNIYIPTASDN--RGSILT-----INHEAGYNSYLFINGDEKVVSYQYKKS 625
DB 652 QOTSXDIYLPSSGWIYARGNAITGGQTIYRSVNPDTLTDMLFI-----KKA 700
QY 626 VSDGQFWERDVVDTRARKPEQGVPTTLVG--YYDPEGTLSSLIYIPAMYGAVGFTYS 683
DB 701 IIPTQ--KVQDYVGQASVTSVDPVFPDTTQSSFTIYDDG-----ASYNYESGTYF 750
QY 684 DDSQNLSDNDQLOVDTKEGQLFRLANHRANNTVMNKFHINVPTESQPTQATLVCNKI 743
DB 751 KQWMTAQDNG-----SGSLSFILGAKSGSYTPALQSYI---VKLHCSAGTSTVNS- 798
QY 744 LDTKSLTAPPEGLTYVNGQALPAKENEGCIVSVNSGKRYCLPVGQORSYSLPDMIVGOE 803
DB 799 -----AAWTSVASLEALKAAAGG-----WATGKD 823
QY 804 VYVD 807
II

```

Db      824 IYGD 827
RESULT 14
Q912M3 PRELIMINARY; PRT; 2468 AA.
AC
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PA1874.
GN PA1874.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales.
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=2043737; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber K.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Wu A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Ku Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA *Complete genome sequence of Pseudomonas aeruginosa PA01, an
RA opportunistic pathogen.*;
RL Nature 406:959-964(2000).
DR EMBL; AE004613; AAG05263.1; -
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR006162; Psantne_attach.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2468 AA; 284414 MW; 13596AFAB2C4B899 CRC64;

Query Match 3.2%; Score 148.5; DB 16; Length 2468;
Best Local Similarity 19.5%; Pred. No. 2.3;
Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 10 NTSOPINDLQSLAAEVKFAQSILP-----AHP-----KEGDSQPH----- 46
Db 324 NSTSPVTVEADTTAPAPATDVQVAPDGGSSVTGNAEPGATVGVDTDGQDPTTVVVGPG 383
QY 47 -----LTSLRKSLLLVRPVKADKTPQVEARD-----DNNKIL----- 80
Db 384 GSFEVPLNPLTNGETVTVITVDPAAGSSPTVTAEPDPPAPQVNASNGSVLSGTAEAG 443
QY 81 -----GTLTLXPSSLPD-----TIYHLDGV 101
Db 444 VTIVITDGNPNIGOTSADANGNSFTPGSOLPDGTVVNVVARDAAAGNSSPATSIITDVG 503
QY 102 PEGGIDFTPHNGTKKIITVAEVNK---LSDASGSIHSHLTNVALVEIHTANGRWYRDI 158
Db 504 APNAPVVEFSGSE---LSGTAEPGSSVTLTDGNGNPICQTAD-----ANGNW---S 550
QY 159 YLPQGDPLEGKWRVFS---SAGYSS-----TFYGDGRKVTLSVGNLTLLFKYVNGOWF 208
Db 551 FTPSTPLPDGIVNVVARDAAAGNSPPASVTVDAVAPATVDPDSNGTTL----- 600
QY 209 RSGELENRRITIAQHIWSAELPAHWIVPLNLVILKQGLNSGLRDLKIGAPGELLHTID 268
Db 601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 269 IGMFTPRDRFDFAKDKFAHREYFOTIPVSRMIVNNYAPLHLKVMPLTGLLTD----- 323
Db 620 IQGVTA-----DGSNGWTFETSTPLPNCVTVNATATD-----PSGNASSPASVTV 664
QY 324 -----MDPGNGWISHTMRQRIKELVSHGIDNANYGLNSTAGLGENSHPYVVAQL 374
Db 665 DAVAPATPVVNPNSNGTTLSTG-----AEPGATVTLTDG-NGNP---IQGV 705

```

```

QY 375 AAHNSRGNTA-----NGIQVH-----GGSGGGGIVTLDL-----TL 405
Db 706 TADGS-GNMSFTPTPLPNCVTVNATATDASGNTSAGSSVTVDVAPATPVINPSNGTTL 764
QY 406 GNEFSHEVGHNYGLHYVDGFGKSVHRSAAENNSNTHGWDGDKKRFIPNEVPS-QTNEKSC 464
Db 765 SG--TAEPGSSVTL---TDGNGNPICQVTDAGSGNW-----SFTSPSTPLADGTV 808
QY 465 LNNOCQEPDGHKFGFDAMAGGSPFSAANRFTMYTPNSSAIIOQFFENKAVFDSRSSTGF 524
Db 809 VNATATDP-----AGNTSGOGSTTVDCVAPTPTV-----NLNCGSSLSGTA- 850
QY 525 SKWNADTQEMEPYEHTIDRAEQITASVNSLSESKMAELMAEYVYVHVMNGNWRNIIY 584
Db 851 -----EPGS-----TVILTDCNGNPVIAEVTAD-----GSGNWT---VT 880
QY 585 PTASADNRGSIILTNHEAGYNSYLFINGDEKVVYSGYKKSFSVSDGQFVKERDWDTREAR 644
Db 881 PSTPIANGTVVNVVQAQAAGNS-----SPGASVT-----VDSQAPA 916
QY 645 KPEQFGVPVTVTLVGYDPEGTLSYIYPAMYGYFTYSDSDSONLSDNDCQLQVDTKEGQ 704
Db 917 APVNVNSNGTTLSTGTAEPGATV-----TLTDCNGNPIC---QVTDG-GSGN 958
QY 705 LRFLANHRANNTVMKHFHINVPTESQPTOATLVCKNKKILDTKSLTPAPEGLTYVNGQA 764
Db 959 WSFTPGTPLANGTVVN-----ATASDPTGNTSAPASTTVD--SVAPAAP-VVNPSNGAE 1009
QY 765 LPAKENECCIVSVNSGKRYCLPYGQBSGYSLPDW 798
Db 1010 ISGTAEPGATVTLTDGSGN--PIGQVTDAGSGNW 1041

RESULT 15
Q8KRR3 PRELIMINARY; PRT; 3692 AA.
AC Q8KRR3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein.
OS Fusobacterium nucleatum.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=851;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PK1594;
RA Ma X., Kempf M.J., Jewett A., Park H.-H., Shi W.;
RT "Cloning and analysis of Fusobacterium nucleatum apoptosis-inducing
RT genes.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF525505; AAM90995.1; -
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
SQ SEQUENCE 3692 AA; 389826 MW; DFF61C00BA0ED180 CRC64;

Query Match 3.2%; Score 148.5; DB 2; Length 3692;
Best Local Similarity 18.8%; Pred. No. 4.3;
Matches 174; Conservative 106; Mismatches 319; Indels 327; Gaps 39;

QY 48 TSLRKSLLLRVPKADKTPQVEARDNNKIL-----GTLTLXPSSLPD 93
Db 162 SSVPTSLL-----PDKKTIIVSV--NNNRALVWDEARENFDQMKGTINLCNQNMGI 213
QY 94 TIY--HLDGVPEGID-----FTPHNGTKKIINTVAEVNKL-----SDASGS 133
Db 214 DLQCTHIGGSAGKATNPAISTIRNEGVITGH-ATNKYDNKKATKEQVAFGFSNADASN 272
QY 134 SIHSHLTNNALVEIHTANGRWYRDIYLPQGDPLEGKWRVFSVSSAGYSTVYFGDRK-VTL 192
Db 273 ATMTHTMINNEITLNPSSAGIQ--LKPEDP-----FYWDPNMGTL 311

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:20:57 ; Search time 21.7788 Seconds
(without alignments)
3912.317 Million cell updates/sec

Title: US-10-002-309B-2
Perfect score: 4709
Sequence: 1 MKLYLSCITLAPLAIGVFS.....AWNGQYLDKPRSMRVYK 886

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4709	100.0	898	T42131	probable toxR-regu
2	2604.5	55.3	2	T00316	toxR-regulated lip
3	689.5	14.6	1002	T09438	toxR-activated lip
4	689.5	14.6	1013	B82276	ToxR-activated gen
5	671	14.2	1335	A82494	TagA-related prote
6	290	6.2	957	H82261	hemolysin-related
7	166.5	3.5	1217	F97177	alpha-glucosidase
8	161	3.4	1038	JC6027	115K outer membran
9	159.5	3.4	2367	S70172	toxin B - Clostrid
10	156.5	3.3	1157	F97255	fusion of alpha-gl
11	154	3.3	1211	S54500	alpha,alpha-trehal
12	153.5	3.3	2366	S10317	toxin B - Clostrid
13	148.5	3.2	2468	A83412	hypothetical prote
14	148	3.1	1021	I40805	collagenase - Clos
15	147.5	3.1	1530	AH1396	peptidoglycan anch
16	147	3.1	1377	D90538	hypothetical prote
17	147	3.1	1386	AC1533	surface protein (L
18	138.5	2.9	973	B86547	polymorphic outer
19	138.5	2.9	973	F72076	polymorphic outer
20	138.5	2.9	995	C81593	polymorphic membra
21	138.5	2.9	1268	B99789	hemagglutinin/hemo
22	138.5	2.9	1270	E85649	hypothetical prote
23	138	2.9	2894	C64474	hypothetical prote
24	137.5	2.9	2529	B64635	toxin-like outer m
25	136.5	2.9	3705	AD0123	probable autotransp
26	136	2.9	2167	AF1489	cell wall-associat
27	135	2.9	1579	B91490	probable invasiv
28	135	2.9	1700	G86131	probable invasiv 2
29	135	2.9	2051	S34688	enoyl-[acyl-carrie

ALIGNMENTS

RESULT 1
T42131

Probable toxR-regulated lipoprotein tagA - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C:Accession: T42131; T00210
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Esc
A:Reference number: Z22068; MUID:98391744; PMID:9722640
A:Accession: T42131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-898 <BUR>
A:Cross-references: EMBL:AF074613; PIDN:AAC70099; 1
A:Experimental source: strain EDL933; serotype O157:H7
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo
S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemor
A:Reference number: Z14127; MUID:98290540; PMID:9628576
A:Accession: T00210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 569-898 <MAK>
A:Cross-references: EMBL:AB011549; MID:d1204561; PIDN:BAA31757; 1; PID:d1032718
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952
C:Genetics:
A:Genome: plasmid p0157
A:Note: L7031

Query Match	100.0%	Score 4709;	DB 2;	Length 898;
Best Local Similarity	100.0%	Pred. No. 9.3e-297;		
Matches 886;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKLYLSCITLAPLAIGVFSATAADNNSAIYFNTSQPINDLQGLSAAAEVKFAQSQILPAH 6C		
DB	13	MKLYLSCITLAPLAIGVFSATAADNNSAIYFNTSQPINDLQGLSAAAEVKFAQSQILPAH 72		
QY	61	PREGDSQPHLTSLRKSLLLRVPVKADKTPVQVEARDNNKILGTLTLPSPSSLPDTIYH 120		
DB	73	PREGDSQPHLTSLRKSLLLRVPVKADKTPVQVEARDNNKILGTLTLPSPSSLPDTIYH 132		
QY	121	LDGVEGGIDFTPHNCTKKIINTVAEVNKLSDAGSSISHSLTNALVEIHTANGRWVRD 180		
DB	133	LDGVEGGIDFTPHNCTKKIINTVAEVNKLSDAGSSISHSLTNALVEIHTANGRWVRD 192		
QY	181	IYLPQGPDLLEGKMWRFVSAGYSSTVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENR 240		
DB	193	IYLPQGPDLLEGKMWRFVSAGYSSTVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENR 252		
QY	241	ITYAQHNSAELPAHWIYVPLNLVKNQNLGRLNDIKTGAPGELLHTIDIGMLTTPRD 300		

112	QY	SSLPTDIYHLDCVPEGGIDFTPHNGTKKIINTVAEVNKLSDASGSSIHSHLTNNALVEIHT	171
113	QY		172
114	QY		173
115	QY		174
116	QY		175
117	QY		176
118	QY		177
119	QY		178
120	QY		179
121	QY		180
122	QY		181
123	Db	SSLPTDIYHLDCVPEGGIDFTPHNGTKKIINTVAEVNKLSDASGSSIHSHLTNNALVEIHT	182
124	QY		183
125	QY		184
126	QY		185
127	QY		186
128	QY		187
129	QY		188
130	QY		189
131	QY		190
132	QY		191
133	QY		192
134	QY		193
135	QY		194
136	QY		195
137	QY		196
138	QY		197
139	QY		198
140	QY		199
141	QY		200
142	QY		201
143	QY		202
144	QY		203
145	QY		204
146	QY		205
147	QY		206
148	QY		207
149	QY		208
150	QY		209
151	QY		210
152	QY		211
153	QY		212
154	QY		213
155	QY		214
156	QY		215
157	QY		216
158	QY		217
159	QY		218
160	QY		219
161	QY		220
162	QY		221
163	QY		222
164	QY		223
165	QY		224
166	QY		225
167	QY		226
168	QY		227
169	QY		228
170	QY		229
171	QY		230
172	QY	TANGRWVRDIYLPQCPDLGGRWVRFEVSAGYSSTVFYDGRKVTLSVGNLTLLFKFYNGQWF	231
173	QY		232
174	QY		233
175	QY		234
176	QY		235
177	QY		236
178	QY		237
179	QY		238
180	QY		239
181	QY		240
182	QY		241
183	QY		242
184	QY		243
185	QY		244
186	QY		245
187	QY		246
188	QY		247
189	QY		248
190	QY		249
191	QY		250
192	QY		251
193	QY		252
194	QY		253
195	QY		254
196	QY		255
197	QY		256
198	QY		257
199	QY		258
200	QY		259
201	QY		260
202	QY		261
203	QY		262
204	QY		263
205	QY		264
206	QY		265
207	QY		266
208	QY		267
209	QY		268
210	QY		269
211	QY		270
212	QY		271
213	QY		272
214	QY		273
215	QY		274
216	QY		275
217	QY		276
218	QY		277
219	QY		278
220	QY		279
221	QY		280
222	QY		281
223	QY		282
224	QY		283
225	QY		284
226	QY		285
227	QY		286
228	QY		287
229	QY		

RESULT 3
T09438
toxR-activated lipoprotein - Vibrio cholerae
N:Alternate names: toxR-activated gene A protein
C:Species: Vibrio cholerae
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
R:Accession: T09438; A39108
R:Carrollis, D.K.R.; Johnson, J.A.; Bailey, C.C.; Boedeker, E.C.; Kaper, J.B.; 1998
Proc. Natl. Acad. Sci. U.S.A. 95, 3134-3139, 1998
A:title: A Vibrio cholerae pathogenicity island associated with epidemic and p
A:reference number: Z16672; MUID:98169509; PMID:9501228
A:Accession: T09438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1002 <KAR>
A:Cross-references: EMBL:AF034434; NID:g3004923; PIDN:AAC12274.1; PID:g3004926
R:Parson, C.; Tauxem, E.; Mekalanos, J.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 1641-1645, 1991
A:title: ToxR regulates the production of lipoproteins and the expression of s
A:reference number: A39108; MUID:91156664; PMID:2000374
A:Accession: A39108
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-25 <PAR>
A:Cross-references: GB:M60658
C:Genetics:
A:Gene: tagA
A:Note: part of the pathogenicity island (VPI); associated with epidemic and p
C:Keywords: lipid binding; lipoprotein

Query Match 14.6%; Score 689.5; DB 2; Length 1002;
Best Local Similarity 24.2%; Pred.No. 2.le-36;
Matches 232; Conservative 151; Mismatches 79; Indels 295; Gaps 39

QY 27 NSAIYFNTSQPINDLQSLAAEVFAQSQILPAHPKPEDSQPHLTSLRKSLLLVPR-VKA 85
+ : : : : | | | | : : : : : : : : : :
Db 137 NEIDYYHTEIEKNKYGVLEGEVFPQTGTVHI--SPEGRKNEPEITIGDALIIFKPSIKN 194

QY 86 DKDTVPVOVEARDNNKKLGTTLTLYPPSSLPDITIYHLQGVPEGGDTPFHNGTKKIINTVA 145

```

Db      195 SSSILMKIYSEDG---LTSKVMKSPSMLPKTDQPID-----ID----- 230
Qy      146 EVNKLSDAGSSSTHSHLTNNALVEIHTANGRWYDIYLPQDPDECKMVFYSSAGYSST 205
Db      231 ----- 230
Qy      206 VFYGDRAKVTLSVGNLTLLFKYVNGQWFRSGELENNR-ITYAQHIWSAELPAHRIWIPGLNLV 264
Db      231 -----ENNVVSYSNSYWSAEIIPWKKMSGMSLH 259
Qy      265 IKOGNISGRL-----NDIKIGAPGELLHTIDIGMLTTPRDRPFAKDKAHR-EYFQTI 318
Db      260 FEDEN--GNLGIIESERIKFSAPSELIIONIDGLMLYKPRGRNIVIKELERTAVDYFOKV 317
Qy      319 PVSRMIVNYPAPLHKEVMLPTCELLTMDPGNGGHSCTMRQIRICKELVSHGIDNANYG 378
Db      318 PVSKLIFSDTPIHFEKITLPNGTVITYEKSADIGGWHQGMREAVKALYSTOINNANLG 377
Qy      379 LNSTAGLGENSEHPYVVAQLAAHNSRGNANGIOVHGSGGGGIVTLDSTLGNFSEHVG 438
Db      378 IVASSGYSO-QYNRLTNHITAHTNIGYNNGVVHGGSGGGIVTLENTLHNEMSHELGH 436
Qy      439 NYGLGHVYDGFSGVHRSANNNSTWGDGDKRFPNFPYPSQTNKESCL--NNQCEPFP 496
Db      437 NYGLGHVYVAG--GTSH---GPDTSWGDGYKRFANFDWKRSPQSNIRPDNQEVVKPF 490
Qy      497 -DGHKGFCDAMAGSSPF--SAANRFMTYTPNSSAIIORFFENKAVFDSRSSTGFSKNAD 553
Db      491 MDKITYLMDAMSGYDHONGIISRYTLHHPYVARIIDMLKNGAVV---INNNDYMYND-E 546
Qy      554 TQEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVVVKHVMNGNWTNRNIYPTASAD 613
Db      547 LKNIIYVYKGT-----NFKVPKKGVPVVTILGVDPDKIN-----PSQLYPTIYS-- 591
Qy      614 NRGSLITIN-----HEAGYN-----SYLFING-- 635
Db      592 NYGNIFOLEKPRSESSLGQYKVDVNYLDRVNTNHTLNVNRKEEKICRFSYLSPRGKK 651
Qy      636 -----DEKVVYSGYKKSFSYSDGOFWKEKROVDVTRFARKPEQFGVPTTLVG---Y 682
Db      652 FEFGLVEDIENKICTGGRSITHLEDG---KKNPI-----ESKYNDYF---LLSIDGGEIS 701
Qy      683 YDPEGLT-----SSYIYPAAMYA-----YGFYSDSDSONLS--- 713
Db      702 YVPDSTIGESKICSLKMSGIVYGAGFIKGNSCQIDGVFMNGFOWAFTLNQSGVNSTIYW 761
Qy      714 DNDQLOVDTKEGQLR-FRLANHRANVTYVANKPHINVTESQPTQATLVNCKNILDTKSL 772
Db      762 SNECVLKIKDKDNKNIIESISIPNYRIEKNQSNKIHNLISREKPIIDINVCGEHELTSIKV 821
Qy      773 TPAPEGLTYTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSG 832
Db      822 SDNPD-----IKLLKGPPII-----VGQEHGYT-----SYEPKLPSPG 852
Qy      833 AKARVLLSDMDNLS-YNRIGFEFVGNVAPADMKYKAMNGQYL-----DFS-KPRSRNVV 884
Db      853 -----WFKHYDNTPEKNEINHELKMRVND-----NDEYICRFSNDSDRMKFV 897

RESULT 4
ToxR-activated gene A protein VC0820 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82276
R:Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, H.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: B82035; MUID:20406833; PMID:10952301
A:Accession: B82276

```

```

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1013 <HET>
A:Cross-references: GB:AE004167; GB:AE003852; NID:9655268; PIDN:AAF93983.1; GSPDB:G
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0820
A:Map position: 1

Query Match      14.6%; Score 689.5; DB 2; Length 1013;
Best Local Similarity 24.2%; Pred.No.2.le-36;
Matches 232; Conservative 151; Mismatches 279; Indels 295; Gaps 39;

```

```

Qy      27 NSAIYFNTSQPINDLOGSLAAEVKFAQSOILPAHPKRGDSQPHLTSLRKSLLLVRP-VKA 85
Db      148 NEIDYHTTEIEKNVYSGLEGEVRFVQTHVI--SPGKNEPEIITGRDALILFSPSINK 205
Qy      86 DKPTVOVARDONNKILGTILYLPSSLPDITIHLDGVPEGIDPTPHNGTKKIINTVA 145
Db      206 SSSILMKIYSEDG---LTSKVMKSPSMLPKTDQPID-----ID----- 241
Qy      146 EVNKLSDAGSSSTHSHLTNNALVEIHTANGRWYDIYLPQDPDECKMVFYSSAGYSST 205
Db      242 ----- 241
Qy      206 VFYGDRAKVTLSVGNLTLLFKYVNGQWFRSGELENNR-ITYAQHIWSAELPAHRIWIPGLNLV 264
Db      242 -----ENNVVSYSNSYWSAEIIPWKKMSGMSLH 270
Qy      265 IKOGNISGRL-----NDIKIGAPGELLHTIDIGMLTTPRDRPFAKDKAHR-EYFQTI 318
Db      271 FEDEN--GNLGIIESERIKFSAPSELIIONIDGLMLYKPRGRNIVIKELERTAVDYFOKV 328
Qy      319 PVSRMIVNYPAPLHKEVMLPTCELLTMDPGNGGHSCTMRQIRICKELVSHGIDNANYG 378
Db      329 PVSKLIFSDTPIHFEKITLPNGTVITYEKSADIGGWHQGMREAVKALYSTOINNANLG 388
Qy      379 LNSTAGLGENSEHPYVVAQLAAHNSRGNANGIOVHGSGGGGIVTLDSTLGNFSEHVG 438
Db      389 IVASSGYSO-QYNRLTNHITAHTNIGYNNGVVHGGSGGGIVTLENTLHNEMSHELGH 447
Qy      439 NYGLGHVYDGFSGVHRSANNNSTWGDGDKRFPNFPYPSQTNKESCL--NNQCEPFP 496
Db      448 NYGLGHVYVAG--GTSH---GPDTSWGDGYKRFANFDWKRSPQSNIRPDNQEVVKPF 501
Qy      497 -DGHKGFCDAMAGSSPF--SAANRFMTYTPNSSAIIORFFENKAVFDSRSSTGFSKNAD 553
Db      502 MDKITYLMDAMSGYDHONGIISRYTLHHPYVARIIDMLKNGAVV---INNNDYMYND-E 557
Qy      554 TQEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVVVKHVMNGNWTNRNIYPTASAD 613
Db      558 LKNIIYVYKGT-----NFKVPKKGVPVVTILGVDPDKIN-----PSQLYPTIYS-- 602
Qy      614 NRGSLITIN-----HEAGYN-----SYLFING-- 635
Db      603 NYGNIFOLEKPRSESSLGQYKVDVNYLDRVNTNHTLNVNRKEEKICRFSYLSPRGKK 662
Qy      636 -----DEKVVYSGYKKSFSYSDGOFWKEKROVDVTRFARKPEQFGVPTTLVG---Y 682
Db      663 FEFGLVEDIENKICTGGRSITHLEDG---KKNPI-----ESKYNDYF---LLSIDGGEIS 712
Qy      683 YDPEGLT-----SSYIYPAAMYA-----YGFYSDSDSONLS--- 713
Db      713 YVPDSTIGESKICSLKMSGIVYGAGFIKGNSCQIDGVFMNGFOWAFTLNQSGVNSTIYW 772
Qy      714 DNDQLOVDTKEGQLR-FRLANHRANVTYVANKPHINVTESQPTQATLVNCKNILDTKSL 772
Db      773 SNECVLKIKDKDNKNIIESISIPNYRIEKNQSNKIHNLISREKPIIDINVCGEHELTSIKV 832
Qy      773 TPAPEGLTYTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSG 832
Db      833 SDNPD-----IKLLKGPPII-----VGQEHGYT-----SYEPKLPSPG 863

```

```
QY 833 AKAYLLSDMDNLS-YNRIGFVGVNVPADMKKKAWNGQYL---DFS-KPRSMRVV 884
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 864 -----WFKHYDNFEKPNINHELKGMRYND-----NDEYICRFNFSDSDRMKFV 908
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
A82494
TagA-related protein VCA0148 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: A82494
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82494
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HEI>
A:Cross-references: GB:AE004356; GB:AE003853; NID:g9657536; PIDN:AAF96061.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0148
A:Map position: 2

Query Match 14.2%; Score 671; DB 2; Length 1335;
Best Local Similarity 25.5%; Pred No. 5.2e-35;
Matches 231; Conservative 108; Mismatches 248; Indels 318; Gaps 36;

QY 39 NDLOGSLAAEYKFAQSILPAHPK--EGDSQPHLTSL--RKSLLVPRVKADDKTPVOVE 94
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 353 NHELGSLEGLSITQHTSVA-PKCNELTGQCHLDAIMNREALLFTPOGEEINQVRAE 411
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 95 ARDDNNKILGTLTPPSLPTDIYHLOGVPEGGIDFTPHNGTKKIIINTVAEVNKLSDAS 154
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 412 VFELDELVQTTLM-LPPSAL-----AASDQPPENGMRKV----- 443
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 GSSISHLTNNALVEIHTANGRWVRDYLPGQPDLEGRKMWRFVSSAGYSTVFYGDRAVT 214
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 444 ----VFSLA----- 449
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 215 LSVGNLTLLFKYVNGWFRSGELENNRITYAQHINSAELPAHWIVPLNLVI-----KQGN 269
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 450 -----WSLPQDWMKPGLSRLTLDNLGREGV 476
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 270 LSGRLNDIKIGAPGELLHTIDIGLTTPRDFDAKD-KEAHREYFOTIPVSRMIVNNY 328
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 477 LS--QGEIQFGGAPELVQNDIGMLPPRDRNTMIONLPTLAADYFOKIPASKLVHADY 534
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 329 APHLKEVMLPTGELLTMDPCNGCWHSCTMRQICKELVSHGIDNANYGLNSTAGLAGEN 388
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 535 TPAHFVPTMPGVVYTDKASGTGWHSGDMREAIKAWSTGINNANVGISSAGYSQO 594
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 389 -----SHPVYAQLAANHSRGYAN-----GIQVHGGSGGGIVTLDTLGNESFHEVG 437
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 595 YNRFRNH-----ITAHTNVGIYTKKDTLDPQVVVHGGSGGGIVTLEATTGNESHGELG 648
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 438 HNYGLGHYDGFSGVHRSANNNTWGDGKKRFPNF-----YPSQTEKSLCLNQ 491
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 649 HNYGLGHM--PYNASIH-----DLESGWGDAPFQRFIGNLHWKGVYTOQQGD-----D 696
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 492 COEPP-DGHKFGFDAMAGSP--FSAANFTWTPNSSLIOFFENKAVFDSRSSTGFS 548
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 697 IVPFFKDAFFRLDAQNGEQEVGTISRTLEHPAQSKRAQRWANGENLDSHSPSGV 756
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 549 KYNADTOEMPEYEHITDRAEQI-----TASVNELSES 580
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 757 QWDOETQRYKAVETDPKPOQVCPVVTLLGIYDPQENPSQIYPLVSYNYGVFELPQP 816
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 581 KMAELMAEYAVVKVHMWNGNWTNIIYIPTASADNRGSLTINHEAGYNSY--LFINGDEK 638
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Query Match	3.5%	Score	166.5	DB 2	Length	1217				
Best Local Similarity	18.6%	Pred. No.	0.021							
Matches	179	Conservative	121	Mismatches	319	Indels	343	Gaps	48	
QY	79	LVRPVKA	---	DDKTPVQV	EARDNNKILGTLTLTPPSSLP	---	DTIYHLDCGVPEGGI	129		
DB	29	IIRPVSAKA	PNMKTISKIKV	ENARV	---	---	SNLSAKLNGDGLQIVNGLD	ETDI	79	
QY	130	---	---	DFPHNGT	KKIINTYVAEVNKLSDASGSIHSLTNNAIV	---	---	EIHTA	173	
DB	80	KICEPOVLK	VDDYKP	-SCQSSSDTLVDP	PNKIWN-TGNIISDLSNDSPWIT	---	---	QKMTIKIS	137	
QY	174	NGRWRVDI	YLPQGDLEGKMY	-RFVSSAGYSSTVFY	GDGRKVTLSVGNLTLLFFKVVNGQW	---	---	---	230	
DB	138	KSDLTMSVY	---	---	DSTGKQIVQQS	TASKSVFTHNSGDR	---	---	FGINGYN	181
QY	231	FR	---	SGELFN	-NRITYAQHWSAELPAHIVPGLNLVIKQGNLSGR	---	---	---	276	
DB	182	FKEDSNKGM	LNRNGTESYI	YAGYQCHGSPFVMS	NDGYGLLVDS	DGSGFTIGDTS	---	---	LOYSGIS	241
QY	277	---	---	IKGARGELL	LHIDI	---	---	---	---	313
DB	242	KTDTYILM	GSPEKIVESD	SGKAPMFPK	WATGFTNTQGWNNIS	---	---	---	SGTGND	301
QY	314	YFOT	---	IPVSRMIY	---	---	---	---	---	356
DB	302	VLNTYRSKQ	IPIDNFC	LDFEKKWGDQNY	GEFFKWTNDF	PAONGOLKAYMD	---	---	SKGLKMT	360
QY	357	GTRQRRI	---	---	GKELVSHG	---	---	---	---	393
DB	361	GIMKPRILAD	SEQARYYTSK	GMWLPDGS	SAADYCSGKMMNVNFAISDV	---	---	---	---	409
QY	394	VAQLAAHNS	RGNYANGTOVHG	SGGGGIVL	DTLSTLGNFEFSHEVGHNYCLGHV	---	---	---	---	445
DB	410	---	---	---	---	---	---	---	---	450
QY	446	VDGFKGS	HRSAENNNSTWG	---	---	---	---	---	---	485
DB	451	IYDG	---	---	QRAYKNQVMSLNR	NYIYAGQYCYGMKSGD	---	---	---	500
QY	486	SCLNQCQ	PEPFDGKHG	FGFD--AMAGSP	---	---	---	---	---	538
DB	501	LSAVN	---	---	---	---	---	---	---	550
QY	539	---	---	FDPSSTG	FSKWNADTQEMPEYHT	IDRAEQITA	---	---	---	581
DB	551	RYPWAFST	AEAAKAKMOLRY	TLIPYIYSDRSAS	QSGGLVRLSLMWEY	PNDSNA	---	---	---	610
QY	582	MAELMAEY	AVYKVHWNGN	TRNTIYP	---	---	---	---	---	630
DB	611	EAMFEGD	MLVSPVYVNGQTS	KSIVLPGN	IDVTTGREY	TGG-OTINYAVDST	---	---	---	669
QY	631	LFING	---	---	---	---	---	---	---	658
DB	670	LFITKSGAI	PIPTQDFENY	YGEKKTIDVYVAD	PPSKATTFDYDDG	TSYDENG	---	---	---	729
QY	659	DVVDTREARK	PEQGVPTTLV	GYDP	EGT	LSYYI	---	---	---	695
DB	730	MTLQTS	SDSKSVQ	ENIDKNT	--GSTPD--	---	---	---	---	785
QY	696	AMYGAYG	FTYSDDSQ	NSLSDN	DCQLQVDT	KEGQLRFLRANHR	---	---	---	755
DB	786	ALKSASGE	GYASGT	---	---	---	---	---	---	826
QY	756	TOATLV	CNNK	---	---	---	---	---	---	808
DB	827	VTITAAAN	PKGTGYG	PTQVTS	LASKSDAT	IYYTL	---	---	---	875

Qy	185	QGPDLGKAVRFVSSAGY--SSTVFTG-----DRKVTLSVGN	220
Db	1597	---NFTQSNIKFTILDANIISGTSIQOEFIDENNIOPIKFTNLTETNYLVGN	1653
Qy	221	L-LFKYVNGOWPFRSELENNRIYAO-HWSAELPAHWIIVPGLNLVIKQGNLSG--	272
Db	1654	QMWIPEPTDLDGSDIISSTVINFOSKYLIGDSCVKNVVISNIIYDEINIIPTVETNN	1713
Qy	273	-----RLNDIKI-----GAPGELLHTIDIGMLTTPRDRF-DFAK	306
Db	1714	TYPEVIVLDANYINEKINVININDLSIRYWSNDGDFILMSTSEENKVSQVKRFPVNFVK	1773
Qy	307	DKE-AHREYF---QTIPYSRMIVNNYAPLHLKEYML-----	338
Db	1774	DKTLANKLSFNFSDKQDVPVSEIIL-SFTPSYYEDGLGIDGLVLSYNEKFIYNNFGMM	1832
Qy	339	-----PTGELLTD-----MDPGNGHWSGTMQRIGIKELVSHGI	372
Db	1833	VSGLIYINDSLYFKPPVNNLIITGVTVGDDKYFFNPINGAAS-----IGETI----	1882
Qy	373	DNANYGLNSTAGIGENSHPHYVVAQLAAHNSRGNY-----ANGIOVHGSGGGIVTLTLDST	427
Db	1883	DDKNYFPNQSG-----VLQTGVFSTEDGFKYPAPAN-----TUDEN	1918
Qy	428	LGNE--FSHEVGHNYGLHYVDGFKGVSHERSAENNSTGWG---DGDKRFFIPNFYPSQ	481
Db	1919	LEGEAIDFTOKLIIIDENIIYFEDNYRGAVE-----WKELDGEMHYFSPE-----	1962
Qy	482	TNEKSCLINOCQBPFDGKHGCFDAMAGGSPFSAANRF--TWYTPNSSAIIQRF--EN	535
Db	1963	-----TGKAFKGLNQIGDDKYFFNSDGVMOKGFPVSINDN	1996
Qy	536	KAVFDSRS--STQFSK-----WNADTOEMPEYHTIDRAEQITASVNELSESOKAELMA	587
Db	1997	KHYFDDSGVMKVGYTEIDGKHFFPAENGEMOIGVFNTEDGFKYFAHNEDLGNEEGEIS	2056
Qy	588	EYAVVKVHMNGWNTRNIIYPTASADNRGSIIITINHEAGVNSYLFIINGDEKVVSGYKKS	647
Db	2057	YSGIL-----NPNKKIYY---FDDSFTAVGVKQLEDGSKYFF---DEDTABAYIGLS	2103
Qy	648	FVSDGQKWERDVVDTREARKEQFG-VPVTTLVGYYPDQECTLSSYIYPAMYGAIFYTS	706
Db	2104	LINDQYFFNDGDI-----MQGVFTINDKVFYFSDSGIIES-----GVONI	2145
Qy	707	DDSONLSDNCQLOV---DTKEGQLRFLANHRANNTVMKFIHNVPTESQTOATLVCN	763
Db	2146	DDNYFYIDDRNGIVQIGVFDTSDGFKYPAPA-----NTVNDNIY-----	2183
Qy	764	NKILDTKSLTAPAE-----GLTYIVNGOALPAKENGECIVSNSGKRYCLPVGORS--GY	816
Db	2184	GOAVEYSGLVRGVDVYFGETYIETGWIYDMENE-----SDKYFFVPETKACKGI	2236
Qy	817	SLPOMIVQOEYVYDSGAKAKVLLSDWNLSY--NRIGEF-VGVNVPAD	861
Db	2237	NLIDDI---KYFDEKGMTRTGLISFNNNYFNGENGEIOFYGINIED	2281

RESULT 10
F97255
fusion of alpha-glucosidase (family 31 glycosyl hydrolase) and glycosidase (Trea/MaIs
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97255
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
R:Accession: F97255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1157 <XUR>
A:Cross-references: GB:AE001437; PIDN:AAK80833.1; PID:q15025938; GSPDB:GN00168


```
Db 392 NVILTEDQPKIIVHKYVIMSTEFNKNKEQ--DNTNIGL-----AKMIA 434
Qy 400 HNSRCNYANGIOVH-----
Db 435 LNSKGNTEKLLSSHKRAWYDIYNDAFIEIPSDSLEMTAKSSLPHLLANTRDYNVSSDRG 494
Qy 414 ---GGSG---GGGIVTLSTLGN-----FSHEVGHNTYGLGHYVDGFKGSVHRSANN 460
Db 495 LPVGVSGLSDSYGGMVFWADIMWEFALLPPFPNVAQ-----MNNRYNATHSQAKLN 548
Qy 461 NSTWGWGDKKRFTPNFYPSOTNEKSLNNOCCPPFGCHKFGFDPAMAGGS-----510
Db 549 AEKYGFGAIYPTWTSKYANTCTGCPGVDEYHINVDVAMASFYILNGHEGIDDEYLRY 608
Qy 511 --PF--SAANRFTMYTP--NSSAIQRFEE-----NKAVFDSRSSTGFSKNAD 553
Db 609 TTWFIKNAOFFYAYKYNSLSGLYETYNLTDPDEFANHINNAFTNAGIKTLLKWATD 668
Qy 554 TQEMEPYEHTIDRAEQITASYNELSESMAELMAEYAVVYHVMWNGMWRNIYIPTASAD 613
Db 669 -----IGNHGEVVDPKWSEI-----SKDIYIPRSSN 696
Qy 614 NRGSLITINHEAGYSYL-----PINGDEKVVSGYKSFVSDQFVKER 658
Db 697 -----ITLEY--SGMNSSVEIKOADVTLMVYPLGYIN--DESILNNAIK-----DLYYYSER 744
Qy 659 DVDVTREARKPEQ--FGVDPVTLVGYDPEGTLSYVYPMYCAYGFTYSDSDSNLSDN 715
Db 745 -----QSASGPMYYPFVAAAGLLNHGSSQSYLYKSVLYPLRAPFAQFSEQSDDN 797

RESULT 12
S10317
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S10317; S21894; S22434
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.
Nucleic Acids Res. 18, 4004, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.
A:Reference number: S10317; MUID:90326540; PMID:2374729
A:Accession: S10317
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-2366 <BAR>
A:Cross-references: EMBL:X53138; NID:g40442; PIDN:CAA37298.1; PID:g40443
R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
submitted to the EMBL Data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21894
A:Molecule type: DNA
A:Residues: 1271-2366 <EIC>
A:Cross-references: EMBL:X60984; NID:g40445; PIDN:CAA43299.1; PID:g40446
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22434
A:Molecule type: DNA
A:Residues: 1791-2366 <VON>
A:Cross-references: EMBL:X60984
C:Genetics:
A:Gene: toxB
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin
```

Query Match 3.3%; Score 153.5; DB 2; Length 2366;

Best Local Similarity 18.6%; Pred. No. 0.41;

Matches 186; Conservative 155; Mismatches 321; Indels 339; Gaps 51;

Qy 27 NSAIYFNTSQPINLQGLAAEYVKAQSIPLPAHPKEGDSQPHLTSLRKSLLLVRPVKAD 86

```
Db 1453 NSELQKNIPYSFVDSGK-----ENGFSNGSTKEG-----LFSVSELPDVLISKVMD 1500
Qy 87 DKTP-----VQVEARDONDKILG-----TLTLYPPSSLPDIYHLD-- 122
Db 1501 DSKPSFGYYSNNLKDVKITKDNVNILTYGLYKDDIKISLSTLQDEKTIKLSVHLD 1560
Qy 123 GVPEGGIDFTPHNGTKIINTVAEVNKLSDASGSSHSHTLNALVEIHTANGRWYRD-- 180
Db 1561 GVAE-----ILKFMNPKGNTN--TSDSLASFLESNNIKSIFVNFLOSNIKFIIDAN 1609
Qy 181 -----IYLPQGPDLGKMAVRFYSSAGYSTVEYDGRKVTL-----S 216
Db 1610 FIISGTTISIGQFECIDENDNIQPIFKF--NTLETNYTLVGNRQNMVPEPNYDLDSDG 1668
Qy 217 VGNLTL---PKYVNGQWFRSGELEN--NRITYAQRHWSAEL---PAH-----WIVPG 260
Db 1669 ISSVTINFSQIYLG-----IDSCVNKYVISPNIYDTEINIPVYETNNYTPYEVILVD 1721
Qy 261 LNLVIKQGNLSGRINDIKI-----GAPGELLHTIDIGMLTTPRORF--DFAKDKE--AHR 312
Db 1722 ANYINEKINYN--INDLSIRYVMSNDGNDFILASTSEENKVSQVKIRFVNVFKDKTLANK 1779
Qy 313 EYP-----QTIPIVSRMIVNNYAPLHLKEYML-----336
Db 1780 LSFNFSQKQDVPVSEIIL--SFTPSYTEDGLIGYDLGLVSLYNEKFYINNFMMVSGLIYI 1838
Qy 339 -----PTGELLTD-----MDPCNGSWHSGTMRQIRIGKELVSHGIDNANYGL 379
Db 1839 NDSLYYKPPVNNLITGFTVVGDDKYFFNPINGAAS-----IGETI-----IDDRNYF 1888
Qy 380 NSTAGLGENSBPYVVAQAAAHNSRGNY-----ANGIQVHGSGGGGIVTLDSTIGNE-- 431
Db 1889 NQSG-----VLQTGVFSTEDGPKYPAPAN-----TLDENLEGEAID 1924
Qy 432 FSEVGHNYGLHYVDGFKGSVHRSANNSTWGW---DQKKRFPINFPSTQNEKSCL 488
Db 1925 FTGLKIITDENIYFDNDYRGAVE-----WKELDGEMHYFSPE-----1961
Qy 489 NNQCQEPFDGKHGFGDAMAGGSPESAARF--TWYTPNSSAITORFF---ENKAVFDSR 542
Db 1962 -----TGKAFKGLNQIGDYKYFYNSDGVKQKGFVSINDNKHFFDSDS 2002
Qy 543 S--STGFSK-----WNADTQEMEPYEHTIDRAEQITASYNELSESMAELMAEYAVKY 594
Db 2003 GVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHNEDLGNEEGEISYSGIL-- 2060
Qy 595 HMWNGNTRNIYIPTASADNRGSLITINHEAGYSNLYFINGDEKVVSGYKSFVSDQGF 654
Db 2061 -----NFNKKIYY---FDDSFYAVGVKDLSDGSKYYF---DEDTAEAYIGLSLINDQY 2109
Qy 655 WKERDVYDTRARKPEQFG--VPVVTLLVGYDPEGTLSYIYPAMYGAIFYSDSDSNLS 713
Db 2110 YFNDDGI-----MQGVFTINDKYFYSDSGIIES-----GVQINDDNYFYI 2151
Qy 714 DNDQQLQV---DTKEGQLRFLANHRANNTVMKFNHINVTESOPTOATLVCNKILDTK 770
Db 2152 DDNGIVQIGVFDTSQYKYPAPA-----NTVNDNIY-----GQAVEYS 2189
Qy 771 SLTPAPE-----GLTYVYNGQALPAKENEGCIVSVNSGKRYCLPVGORS--GYSLPDWIV 823
Db 2190 GLVRVGEDVYFYGETTYTETGIWYDMENE-----SDKYTFNFPETKACKGINLIDDI- 2241
Qy 824 GQEVYVDSGAKAKVLLSDMDNLSY--NRIGEF--VGNVNPAD 861
Db 2242 --KYIFDEKGMRTGLISFENNYYFNENGEMQFYGINIED 2280
```

RESULT 13

A83412

hypothetical protein PA1874 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83412
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoquchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; NID:99947856; PIDN:AAG05263.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874

Query Match 3.2%; Score 148.5; DB 2; Length 2468;
Best Local Similarity 19.5%; Pred. No. 0.92;
Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 33 NTSOPINDLQGSLLAEVKEFAQSQILP-----AHP-----KEGDSOPH----- 69
DB 324 NSTPTVTEAPOTTAPATPDVQVAPDGSSTVTCNAEPGATVGVDTGDCQPTTVVVGPG 383
QY 70 -----LTSRLSKLLLRPVKADKTPVQVEARD-----DNKKIL----- 103
DB 384 GSFEVPLNPLTNGETVTIVTDPAGNSSTPTVTAEPDPPAPQVNASGSLVSGTAEAG 443
QY 104 -----GTLTLYPPSSLPD-----TIYHLQGV 124
DB 444 VTIVITDGNPNIGOTSADANGMSPTGSQLPDGTVVNVVARDAGNSPATSITVDGV 503
QY 125 PEGGIDFTPHNGPKKIINTVAENK---LSDASGSIHSHLTNNALVEIHTANGVRVDI 181
DB 504 APNAPVPSNSE---LSGTAEPGSSVTLTDGNGNPGTQTTAD-----ANGN---S 550
QY 182 YLPQGDLEKMYRFVS---SAGYSS-----TFYGDRTKVTLSVNTLLFYKYNQWF 231
DB 551 FTPSTPLPDGTVVNVVARDAGNSPASPASVTVDVAPATPTVDPSSNGTTL----- 600
QY 232 RSGELENNRITYAQHTWSAELPAHWVPGVNLVKGSLNLSGRINDIKIGAPCELLHTID 291
DB 601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 292 IGLMTTPDRFOPAKKEAHREYFOTIPVSRMTVNNYAPLHLKEVMLPTGELLTD----- 346
DB 620 IGVTA-----DGSNGWTFPTPLPNCVTVNATD-----PSGNASSPASVIV 664
QY 347 -----MDPONGGWHGTHQRIGKELVSHGIDNANYGLNSTAGLGENSEHPYVAQL 397
DB 665 DAVAPATPVVNSNGTTLST-----AEPGATVTLTDG-NGNP--IGQV 705
QY 398 AAHNSRGNTA-----NCIOVH-----GGSGGGIVTLDS-----TL 428
DB 706 TAGDS-GNMSFTPTPLPNCVTVNATDASGNTSAGSSVTVDSVAPATPVINPNSGTTL 764
QY 429 GNEFSHEVGHNYGLHYDFKGSVHRSAENNNSTWGDGKKRFIPNFYPS-OTNEKSC 487
DB 765 SG--TAEPGSSVTL---TDGNGNPIQVTDAGSGNW-----STPTPLADGTV 808
QY 488 LNNQCOPEPDGHKFGDAMAGSGSPSAANRFTMYTPNSSAIIOFFENKAVFDSRSSTGF 547
DB 809 VNATATDP-----AGNTSGGSGTVDGVAPTPTV-----NLSNGSSLSGTA- 850
QY 548 SKWNADTQEMPEYHTIDRAEQITASVNELSEKMAELMAEYAVVVMHNGWNTNIIYI 607
DB 851 -----EPGS-----TVILTDGNGNPIAEVTAD-----GSGNWT---YT 880
QY 608 PTASADNRGSILLITINHEAGVNSLYFINGDEKVVSOQYKKSFVSDGQFWKERDVTDTREAR 667
DB 881 PSTPIANGTVVNVVAQAAGNS-----SPGASVT-----VDSQAPA 916
QY 668 KPQFGVPVTLVGYDPEGTLSSTIYPAMYGYFTYSDDSNLSDNDQQLQVDTKEGQ 727

DB 917 APVWNPSNGTTLSGTAEPGATV-----TLTDGNGNPIG---QVTAD-GSGN 958
QY 728 LRFLANHRANNVTYKPHINVPTESOPTQATLVCKNKKILDTKSLTAPAGLTYTVNQQA 787
DB 959 WSFTPTGTPLANGTVVN-----ATASDPTGNTSPASTTVD--SVAPAAP-VVNPSPNGAE 1009
QY 788 LPAKENEGCIVSVNSGKRYCLFVPGQSGYSLPDW 821
DB 1010 ISGTAEPGATVTLTDGSGN--PIGQVTDAGSGNW 1041

RESULT 14
I40805
collagenase - Clostridium histolyticum
C:Species: Clostridium histolyticum
C:Date: 16-Aug-1995 #sequence_revision 16-Aug-1996 #text_change 21-Jul-2000
C:Accession: I40805; T44355
R:Yoshihara, K.; Matsushita, O.; Minami, J.; Okabe, A.
J. Bacteriol. 176, 6489-6496, 1994
A:Title: Cloning and nucleotide sequence analysis of the colH gene from Clostridium
A:Reference number: I40805; MUID:95050206; PMID:7961400
A:Accession: I40805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1021 <RES>
A:Cross-references: GB:D29981; NID:9563954; PIDN:BAA06251.1; PID:g710023
R:Matsushita, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.
J. Bacteriol. 181, 923-933, 1999
A:Title: Gene duplication and multiplicity of collagenases in Clostridium histolytic
A:Reference number: 222752; MUID:99121032; PMID:9922257
A:Accession: T44355
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1021 <MAT>
A:Cross-references: EMBL:AB014075; NID:93868863; PIDN:BAA34542.1; PID:g3892646
A:Experimental source: strain JCM 1403
C:Genetics:
A:Gene: colH
C:Superfamily: microbial collagenase

Query Match 3.1%; Score 148; DB 2; Length 1021;
Best Local Similarity 18.6%; Pred. No. 0.25;
Matches 169; Conservative 121; Mismatches 330; Indels 288; Gaps 46;

QY 97 DNNKILGTLYPPSSLPDTIYHLDGVPGEIGDTPHNGTKKIINTVAENKLSASGS 156
DB 263 DNNNSWI-----IDNGIYHI--APLGL-----HSNNKIGIETLTVMK----- 299
QY 157 SIHSHLTNNALVEIHTANGVRVDIYLPQGPDLGKQW---RFVSSAGYS---STVFYGD 210
DB 300 -VVPYLSMQ-----HLOSADQIKRHY--DSKDAENKIPLDKPKKGEKCYKPTVTFDD 351
QY 211 RVTLSVSGNTL-----LFKYVNGQWFRSGELENRITYAQHINSAELPAHWIVPG 260
DB 352 GVVIKAGARVEEYKRYLWASKEYNSQFFR-----VYG 386
QY 261 LNLVKGSLNLSGRINDIKIGAPCELLHTI-----DIGMLTTPDRDFADKKEAHR 313
DB 387 IKPLEEGPDDILLNVYINSPEEYKLSVLGYDTNNGMYIEPGTF-FYIERAQES 445
QY 314 YFOTIPVSRMIVNY-----APLHLKEVMLPTGELLTMDPONGGWHSGTMRQRIKEL 367
DB 446 TTYLEELFRHEYTHYLOGRYAVPGWGRTKLYDNDRLTWEEG-----GAL 492
QY 368 VSHGIDNANYGLNSTAGLGENSEHPYVAQLAAHNSGN--YANGIQVHGGSGGGIVTLD 425
DB 493 FA-----GSTRSGILPRKS--IVSNI--HNTTRNNRYKLSDTVH----- 528
QY 426 STLGNFSEVGHNYGLGHYVD-----GFGSVHRSANNNSTWGDGKKRFIPNRY 478
DB 529 SKYGASFEF---YNTAC-MFMDYTNKMDGILNKLMDLAKNNDVD-GYDN-----Y 574

[illegible]

Search completed: October 8, 2003, 16:28:15
Job time : 27.7788 secs

```
QY      8/3 FSKPMSRKY 884
      :| : :|
Db      1438 AEQPNFQIV 1447

Search completed: October 8, 2003, 16:28:15
Job time : 27.7788 secs
```


XX PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX DR Welch RA, Latham WH;
 XX PI N-PSDB; AAD38751.
 XX WPI; 2002-471441/50.
 XX DR N-PSDB; AAD38751.
 XX PT New p0157 plasmid-specified polypeptide found in Escherichia coli and
 XX PT other enterohemorrhagic Escherichia coli, that binds to and cleaves
 XX PT Cl-esterase inhibitor, useful for diagnosing and treating colitis -
 XX PS Claim 1; Page 52-56; 58pp; English.
 XX CC The present invention relates to novel p0157 plasmid-specified proteins
 XX CC found in Escherichia coli EDL933 and other enterohemorrhagic E. coli,
 XX CC designated Stx, that bind to and cleave Cl-esterase inhibitor. Sequences
 XX CC of the invention are useful for diagnosing, preventing or treating
 XX CC haemolytic uraemic syndrome or colitis in a subject infected with an
 XX CC enterohemorrhagic pathogen expressing inhibitor protein. They are
 XX CC useful for testing a molecule for the ability to reduce proteolysis
 XX CC of Cl esterase inhibitor by inhibitor protein. The present sequence
 XX CC is E. coli O157:H7 Stx protein encoded by p0157 plasmid DNA.
 XX SQ Sequence 886 AA;
 Query Match 100.0%; Score 4709; DB 23; Length 886;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 886; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLYLSCTILAPLAIGVFSATAADNNSAIVNTSOPINDLOGSLAAEVEKFAQSOILPAH 60
 DB 1 MKLYLSCTILAPLAIGVFSATAADNNSAIVNTSOPINDLOGSLAAEVEKFAQSOILPAH 60
 QY 61 PREGSQPHLSRLSKLLVPRVADKTPQVVEARDNNKILGTLTLYPPSSLPDTIYH 120
 DB 61 PREGSQPHLSRLSKLLVPRVADKTPQVVEARDNNKILGTLTLYPPSSLPDTIYH 120
 QY 121 LDGVPPEGIDFTPHNGTKKIINTVAEYVKLSADGSSIRHSLTNALVEIHTANGRWYRD 180
 DB 121 LDGVPPEGIDFTPHNGTKKIINTVAEYVKLSADGSSIRHSLTNALVEIHTANGRWYRD 180
 QY 181 IYLPQGPDLGKRWFRVFSAGSYSTVFYGDARKVTLVSGNTLLFKYVQWFRSGELENNR 240
 DB 181 IYLPQGPDLGKRWFRVFSAGSYSTVFYGDARKVTLVSGNTLLFKYVQWFRSGELENNR 240
 QY 241 IYAOHIWSAELPAHWIYVGLNVLTKQNLGRLNDIKIGAPGELLHTIDIGMLTTPRD 300
 DB 241 IYAOHIWSAELPAHWIYVGLNVLTKQNLGRLNDIKIGAPGELLHTIDIGMLTTPRD 300
 QY 301 RFDFAKDKEAHREYFOTIPVSRMIVNNYAPLHLKEVMLPTGELLTMDPGNGWHSGTMR 360
 DB 301 RFDFAKDKEAHREYFOTIPVSRMIVNNYAPLHLKEVMLPTGELLTMDPGNGWHSGTMR 360
 QY 361 QRIGELVSHGIDNANYGLNSTAGLGENSHPYVVAQLAAHNSRGNANGIQVHSGSGGG 420
 DB 361 QRIGELVSHGIDNANYGLNSTAGLGENSHPYVVAQLAAHNSRGNANGIQVHSGSGGG 420
 QY 421 IVTLDTLGLNEFSHEVGHNYGLGHVYDGFKGSVHRSANNSTWGDGKRFPINFPYS 480
 DB 421 IVTLDTLGLNEFSHEVGHNYGLGHVYDGFKGSVHRSANNSTWGDGKRFPINFPYS 480
 QY 481 QNEXSCLNQCQEPDGHGKFGFDAMAGSPESAANRFTMTYTPNSSAIIQRFENKAYFD 540
 DB 481 QNEXSCLNQCQEPDGHGKFGFDAMAGSPESAANRFTMTYTPNSSAIIQRFENKAYFD 540
 QY 541 SRSSGFGSKWADTQEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVYKVMWNGN 600
 DB 541 SRSSGFGSKWADTQEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVYKVMWNGN 600
 QY 601 WTRNIIPTASADNRGSLITINHEAGYNSYLFINGDEKVSQYKKSFSVDSQFQWKRVDV 660
 DB 601 WTRNIIPTASADNRGSLITINHEAGYNSYLFINGDEKVSQYKKSFSVDSQFQWKRVDV 660

QY 661 VDTREARKPEOFGVPVTVTLVGYDDPGTLLSSYIYPAMYGAYGFTYSDDSQNLSDNDCQLQ 720
 DB 661 VDTREARKPEOFGVPVTVTLVGYDDPGTLLSSYIYPAMYGAYGFTYSDDSQNLSDNDCQLQ 720
 QY 721 VDTKEGQLRFLANHRANNTVNNKFHINVPTESQPTQATLVCNNKILDTKSLTPAPEGLT 780
 DB 721 VDTKEGQLRFLANHRANNTVNNKFHINVPTESQPTQATLVCNNKILDTKSLTPAPEGLT 780
 QY 781 YTVNGALPAKENEGGCIIVSVNSGKRYCLPVGQBSGYSLPDWIVGQBYVYDSGAKAKVLLS 840
 DB 781 YTVNGALPAKENEGGCIIVSVNSGKRYCLPVGQBSGYSLPDWIVGQBYVYDSGAKAKVLLS 840
 QY 841 DWDNLSYNRIGEPVGNVNPADMKKVKAWNGQYLDKSKPRSMRVYK 886
 DB 841 DWDNLSYNRIGEPVGNVNPADMKKVKAWNGQYLDKSKPRSMRVYK 886
 RESULT 2
 AAR95011
 ID AAR95011 standard; Protein; 2366 AA.
 XX AC AAR95011;
 XX DT 08-JUL-1996 (first entry)
 XX DE C. difficile toxin B.
 XX KW Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;
 XX KW diarrhoea; therapy; diagnosis; vaccine.
 XX OS Clostridium difficile VPI strain 10463 (ATCC 10463).
 XX PN W09612802-A1.
 XX PD 02-MAY-1996.
 XX PF 23-OCT-1995; 95WO-US13737.
 XX PR 07-JUN-1995; 95US-0480604.
 XX PR 24-OCT-1994; 94US-0329154.
 XX PR 16-MAR-1995; 95US-0405496.
 XX PR 14-APR-1995; 95US-0422711.
 XX PA (OPHI-) OPHIDIAN PHARM INC.
 XX PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
 XX PI Williams JA;
 XX DR WPI; 1996-230603/23.
 XX DR N-PSDB; AAT29247.
 XX PT Fusion proteins comprising non-toxin protein and part of toxin -
 XX PT useful to form anti-toxins against Clostridium botulinum type A, and
 XX PT C. difficile type toxins, and to treat C. difficile intoxication,
 XX PT partic. diarrhoea
 XX PS Claim 36; Page 313-323; 434pp; English.
 XX CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the
 XX CC product of the toxin B gene (AAT29247), is a cytotoxin associated
 XX CC with diarrhoeic disease. It can be obtd. by expression in
 XX CC transformed E. coli hosts of portions of DNA that together cover the
 XX CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,
 XX CC AAR75371-72 and AAR95018), pref. expressed as fusions to polystyrene
 XX CC affinity tags or maltose binding protein, are used to raise avian
 XX CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
 XX SQ Sequence 2366 AA;

Query Match 3.3%; Score 153.5; DB 17; Length 2366;
 Best Local Similarity 18.6%; Pred. No. 0.033;
 Matches 186; Conservative 155; Mismatches 321; Indels 339; Gaps 51;

QY	27	NSAIYFNTSOP	INDLQGS	LAEEYFAQSO	ILPAHPKEG	QSOPHLT	SLRKS	LLLPVKAD	86
Db									
Db	1453	NSELQNI	PYSFV	DECK	-----	ENG	FING	STEG	----
									1500
QY	87	DKTP	-----	VQ	EARDDNN	KILG	-----	TL	LYP
Db									122
Db	1501	DSKPS	FGY	YNNLKD	VKITK	QNVNLT	GYL	KDDIK	ISLSLT
									1560
QY	123	GVPEGG	IDFPH	NGTKKI	INTVA	EVNKL	S	D	SSSTH
Db									180
Db	1561	GVAE	-----	IL	KMRK	GNTN	TS	OSL	MSF
									1609
QY	181	-----	IY	PGDP	LEGR	KYR	V	SS	AGY
Db									216
Db	1610	FIISG	TTSG	EF	IC	DEND	NI	QY	FIK
									1668
QY	217	VGN	TLL	---	FR	YNG	QW	FR	SELEN
Db									260
Db	1669	ISST	Y	NF	SK	LYG	-----	ID	SK
									1721
QY	261	LN	LV	IK	OG	N	LS	GR	LN
Db									312
Db	1722	ANT	INE	KIN	YN	---	IND	US	I
									1779
QY	313	EYF	-----	Q	T	P	S	R	M
Db									338
Db	1780	LS	N	F	S	D	K	O	V
									1838
QY	339	-----	PT	CELL	T	-----	MD	P	C
Db									379
Db	1839	N	D	S	L	Y	F	K	P
									1888
QY	380	N	S	T	A	G	L	E	N
Db									431
Db	1889	N	O	S	G	-----	V	L	T
									1924
QY	432	F	S	H	E	V	G	H	N
Db									488
Db	1925	F	T	G	L	I	D	E	N
									1961
QY	489	N	N	O	C	E	P	D	H
Db									542
Db	1962	-----	T	K	A	F	K	G	L
									2002
QY	543	S	---	S	T	G	F	S	K
Db									594
Db	2003	G	V	M	K	V	G	T	E
									2060
QY	595	H	M	N	G	N	T	R	I
Db									654
Db	2061	-----	N	F	N	N	K	I	Y
									2109
QY	655	W	K	E	R	D	V	D	T
Db									713
Db	2110	Y	F	N	D	D	G	I	-----
									2151
QY	714	D	N	O	C	O	L	O	V
Db									770
Db	2152	D	D	N	I				

РЕЗУЛТАТ

RESULT 3

AAW68388

ID AAW68388 standard; Protein; 2366 AA.

[illegible]

57

QY

Db 1722 ANYINEXINVN--INDLSIRYVWSNCGNDFILMSTSEENKVSQVKIRFVNVFKDKTLANK 1779
 QY 313 EYF-----QTIPYSRMIVNNYAPLHLKEVML----- 338
 Db 1780 LSFNFSKQDVPVSEIL-SFTPSYEDGLIGYDLGLVSLYNKFKYINNFQMVSGLIYI 1838
 QY 339 -----PTGELLTD-----MDPENGWHSQTKMRQIRIGKELVSHGIDNANYGL 379
 Db 1839 NDSLYFKPPVNNLITGFVTVDGDKYFNPINGCAAS-----IGETI-----IDDKNYF 1888
 QY 380 NSTAGLGNSHPYVVAQLAAHNSRGN-----ANGIQVHGGSGGGIVTLDSTLGN----- 431
 Db 1889 NQSG-----VLGTGVFSTDEGKFKYFAPAN-----TIDENLEGAID 1924
 QY 432 PSHEVGHNYGLGHVVDGFKGSHVRSANNSWTGW---DGDKKRFIFNFPYPSQTNEKSCL 488
 Db 1925 FTGLIIDIENIYFDDNYRGAVE-----WKELDGMHFYSP----- 1961
 QY 489 NNQCQEPFDGKHGFDAMAGGSPFSAANRF--TWYTPNSSAIORFF-----ENKAVDSR 542
 Db 1962 -----TKKAPKGLNQIGDYKYFYFNSDGMOKGFVSINDNKHFFDSDS 2002
 QY 543 S--STGFSK-----WNADTQEMPEYEHTIDRAEQITASVNSLSESKMAELMAEYAVVKV 594
 Db 2003 GVMKVGTEIDGKHFFFAENGEMOIGVFNTEDEGFKYFAHNEDLGNEGEISYSGIL-- 2060
 QY 595 HMMNGNTRNIYIPTASADNRGSLTINHEAGYNSYLFINGDERKVSQGYKKSFVSDQGF 654
 Db 2061 -----NFNNKIY---FDSFTAVVGWKLEDGSKYF---DEDTAEAYIGLSLINDGQY 2109
 QY 655 WKERDVTREARKPEFG-VPVTTLVGYDPEGLTSSYIYPAMYKAGFYTSDSONLS 713
 Db 2110 YFNDGDI-----MOVGFYTIINDKYFYFSDSGIIES-----GVONIDDNFYI 2151
 QY 714 DNDQOLQV---DTKEGQLRPLRANRANNTVMKFNHINVPESQPTQATLVCKNKILDTK 770
 Db 2152 DONGIVQIGVEDTSDGKYKYPAPA-----NTVNDNIY-----QCAVEYS 2189
 QY 771 SLTPAPE-----GLTYTVNGOALPAKENEGCIVSVNSKRYCLPVGORS--GYSLPDMIV 823
 Db 2190 GLVRVGEDVYFGETYTIETGWIYDMENE-----SDKYFFNPETKKACKGINLIDDI- 2241
 QY 824 GOEYVDSGAkakVLLSDWDLNLSY--NRIGEF-VGNVNPAD 861
 Db 2242 --KYFDEKIMRTGLISFENNNTYFENGEMQFYINIED 2280
 RESULT 4
 ABG30537
 ID ABG30537 standard; Protein; 1249 AA.
 XX AC ABG30537;
 XX DT 07-OCT-2002 (first entry)
 XX DE Alpha-isomaltosylglucosaccharide synthase #1 mature protein.
 XX KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
 KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
 KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
 KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
 KW cosmetic; drug composition; enzyme.
 XX OS Unidentified.
 XX PN WO200255708-A1.
 XX PD 18-JUL-2002.
 XX PF 09-JAN-2002; 2002WO-JP00052.
 XX PR 12-JAN-2001; 2001JP-0005441.
 XX

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX Kubota M, Maruta K, Yamamoto T, Fukuda S;
 DR WPI: 2002-520129/55.
 DR N-PSDB; ABK88154.
 XX New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
 PT the production of cyclic tetrasaccharide gum -
 PT
 XX Claim 1; Page 101-105; 144pp; Japanese.
 PS
 XX The invention describes novel microbial polypeptides having
 CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
 CC useful for producing a sugar (I) having at least three glucose units
 CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
 CC end. The invention also describes a method for the production of the
 CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
 CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
 CC (I) by treatment of (I) with alpha-isomaltosyltransferase.
 CC (II) and similar sugars in crystalline or syrup form are used as
 CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
 CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
 CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
 CC synthase mature protein.
 XX
 SQ Sequence 1249 AA;
 Query Match 3.24; Score 149; DB 23; Length 1249;
 Best Local Similarity 18.88; Pred. No. 0.027;
 Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;
 QY 105 TLTLYPPSSLPDIYHLDGVPEG--GIDFTPHNGTKKIINTVAEVNKLSDAGSSSIHSH- 161
 Db 19 TLTVDNCAEBSDDLIVQAVQNGILKVDYRPNSTPSAKTPMLDPNKTWSAVCATINTTA 78
 QY 162 -----LTNNALVEI-----HTANGRWROIYLPQGPDLGKMKVRFVSSAGYSTVF 207
 Db 79 NPMTITTSNNKIEITKNPVRMTYKADGTTL--FWPESGGGVFSDGVRFLLHATGDN---M 133
 QY 208 YGRKVTLSVGNLTLLFKYVNGQWFRSGELENNRITVAQHIWSAELPAHIVPGLNLVIKQ 267
 Db 134 YGIRSF-----NAFDSGGDLLRNSNHAH----- 158
 QY 268 GNLSGRLLNIGAPGELLHTIDIGMLTTPRDFPAKDKAEHREYFO--TIPVSRMIV 325
 Db 159 ---AGEQD---SGGPLIWSAGYGLLVSDGYPYTDSTTGQMEFYIYGGTPEGRRYA 211
 QY 326 NNTAPLHL-----KEVMLPTGELLTMDPONGGWHSGTMR-----Q 361
 Db 212 KQVEYYIMLGTPKREIMTDVGE-ITGKPPMLPKHSLGFNFEMDTNOTFTNNVDTYRAK 270
 QY 362 RIGKELVSHGIDNANYGLNSTAGLGNSHPYVVAQLAAHNSRGNVANGIQVHG----- 414
 Db 271 NIPIDAFYDMKKYGETNYGEFAWNTTNPFSASTTSLKSTMD--ARGIKMIGITKPIV 329
 QY 415 -GSGGGGIVT--LDSTLGNFSEHVGHNYGLGHVVDGFKGSHVRSAE---NNNSTWQWDG 468
 Db 330 TKDASANVTTOCTDATNGGYF--YPGHN---EYQDYFIPVTYRSIDIPYNANERAWFNH 383
 QY 469 DK---KRFIPNFYPSQTNKS-----CLNNQCQEPFDGKHGFDAMAGGSP--- 511
 Db 384 STDALNKGIVGWNDETDRKYSSGALYWFNGFTTGHMSQTMDEGGR---AYTSGAQRVW 439
 QY 512 -----FSAANRF--TWYTP-----NSSAIQRFENKAVFDSRSSTGSKWN 551
 Db 440 QTARTFYPGAQRVATTLWSGDIGIQYNGKGERINWAGHQ---EQRAYMLSSVNGQYKWG 496
 QY 552 ADT-----QEMEPY-----EHTIDRA 567
 Db 497 MDTGGFNOQDGTNNPNPDLYARMQFSALTPTFVRVHGNHQQRPWYFGSTAEASKEA 556
 QY 568 EQI-----TASVNELSESKMAELMAEY---AVVK--VDMWN-GNN----- 60-

RESULT 6
 ABP70652
 ID ABP70652 standard; Protein; 1284 AA.
 XX AC
 AC ABP70652;
 XX DT
 DT 06-MAR-2003 (first entry)
 XX DE
 DE Bacillus globisporus protein encoded by SEQ ID 2.
 XX KW
 KW Cyclic tetrasaccharide; sweetening agent.
 XX OS
 OS Bacillus globisporus.
 XX FH
 FH Key Location/Qualifiers
 FT Peptide 1..35
 FT /label= Signal_peptide
 XX PN
 PN WO200272594-A1.
 XX PD
 PD 19-SEP-2002.
 XX PF
 PF 08-MAR-2002; 2002WO-JP02213.
 XX PR
 PR 09-MAR-2001; 2001JP-0067282.
 XX PA
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX PI
 PI Aga H, Higashiyama T, Watanabe H, Sonoda T, Kubota M;
 DR WPI; 2003-092811/08.
 DR N-PSDB; ABV99948.
 XX PT
 PT New cyclic tetrasaccharide compounds useful as sweetening agent for
 PT foods, drinks or pharmaceuticals.
 XX PS
 PS Disclosure; Page 122-129; 133pp; Japanese.
 CC The present invention relates to novel cyclic tetrasaccharide compounds:
 CC cyclo(-6-)-alpha-D-glucopyranosyl-(1-3)-alpha-D-glucopyranosyl-(1-6)-
 CC alpha-D-glucopyranosyl-(1-3)-alpha-D-glucopyranosyl-(1-)- derivatives.
 CC The tetrasaccharide compounds are used as sweetening agents for foods,
 CC drinks or pharmaceuticals. The present protein was used to illustrate the
 CC invention.
 XX SQ
 SQ Sequence 1284 AA:
 Query Match 3.2%; Score 149; DB 24; Length 1284;
 Best Local Similarity 18.8%; Pred. No. 0.028;
 Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;
 QY 105 TLTYPPSSLPDIYHLDGVEG--GIDFTPHNGTKKIINTVAEVNKLSDASGSIHSH- 161
 DB 54 TLTVNGAEPDOLLIVQVQNGILKVDYRNSITPSAKTPLDPNKTWSAVGATINTTA 113
 QY 162 -----LTNNAIVEI-----HTANGWRVDYLPQGPDLGKRWRFVSSAGYSSTVF 207
 DB 114 NPMITTSNMKIEITKPNVRMTVKKADGTTL--FWEPSSGGVFGSDGVFLHATGDN---M 168
 QY 208 YGDRKVTLSVGNLTLFKYVNGQWFSRSELENRTIYAQHIWSAELPAHWIYVPLNLVIKQ 267
 DB 169 YGIRSF-----NAFDSGGDLRNSNHAH----- 193
 QY 268 GNLSGRNDIKIGAPGELLITIDIGMTTPDRDFDAKDEAHREYFQ--TIPVSRMIV 325
 DB 194 ---AGEQGD---SGPLWSTAGVLLVDSGGYPYTDSTTGQEFYGGTPPEGRYA 246
 QY 326 NNYAPLHL-----KEVMLPTGELLTMDPGNGNHSGTWR-----Q 361
 DB 247 KQVEYIYMLGTPKPEIMTDVGE-ITGKPPMLPKWSLGFPMFEWDTNQTEFTNNVDTYRAK 305
 QY 362 RIGKELVSHGIDNANYGLNSTAGLGENSHPPYVAQLAAHNSRGNYANGIQVHG----- 414

Db 306 NIPIDAYAFDYDMKKYGETNYGEFANNTNFPASSTLSLKSTMD-AKGIKIGITKPRIV 364
 QY 415 -GSGGGGIYV--LDSTLGNFEFSHEVGHNYGLHYVDGFGSGVHRSAE---NNNSTWGDG 468
 Db 365 TKDASANVTQTGTDATNGGYF--YPGHN---EYODYFIPVTVRSIDPYNANERAWFWNH 418
 QY 469 DK---KRFIPNFYPSOTNEKS-----CLNNQCOEPDGHKFGFDAMAGGSP--- 511
 Db 419 STDALNKGIVGWNDETDKVSSGGALYWFNGFTTGHMSOTNTEGGR-----ATSGQRVW 474
 QY 512 -----FSAANRF--TWYTP-----NSSAIIOFFENKAVFDSRSSTGFSKNW 551
 Db 475 QTARTFPGAQRYATTLWSDIGIOYNKGERINWAAQM---EQRAVMLSSVYNGQVKGW 531
 QY 552 ADT-----QEMEPY-----EHTIDRA 567
 Db 532 MDTGFGNQDGTNNPNPDLYAHMMQFSALTPTVFRVHGNHQQORQPMYFGSTAEEASKEA 591
 QY 568 EQI-----TASVNELSESKMAELMAEY---AVVK--VHMWN-GNW----- 601
 Db 592 IQLRSLIPYMYAYERSAYENGNGLYRPLMQAYPTDAVKNYTDAWFGDWLLAAPVDR 651
 QY 602 ---TRNIYIPTASADN--RGSILT-----INHEAGYNSYLFINGDEKVVSGYKKSP 648
 Db 652 QQTSKDIYLPSSGWIDYARGNAITGGGTIRYSVNPOTLTDMPLEI-----KKA 700
 QY 649 VSDGQFWKRDVYDTREARKPEQFGVPVTLVG--YDPEGTLSSYIYPAMYGAYGFTYS 706
 Db 701 IIPQ--KVODYVQASVTSVDVDFDTQSSFTYDDG-----ASYNYESGTYP 750
 QY 707 DSDNLSDNDCLQVDTKEGQLRFRLANHRANNTVMNKFHINVTESQPTQATLVCNKI 766
 Db 751 KONMTAQDNG-----SGSLFTLGAKSGSYTPALQSYI---VKLHGSAGTSVTNNS- 798
 QY 767 LDTKSLTPAPEGLTYTVNQALPAKENEGCIVSVNSGKRYCLPVGORSYSLPDWIVGOE 828
 Db 799 -----AAMTSYASLEAKAAAGE-----WATGKD 823
 QY 827 VYVD 830
 Db 824 IYGD 827
 RESULT 7
 ABP57724
 ID ABP57724 standard; Protein; 1284 AA.
 XX AC
 AC ABP57724;
 XX DT
 DT 23-JAN-2003 (first entry)
 XX DE
 DE Protein #1 related to isomaltose production.
 XX KW
 KW Isomaltose.
 XX OS
 OS Unidentified.
 XX PN
 PN WO200288374-A1.
 XX PD
 PD 07-NOV-2002.
 XX PF
 PF 25-APR-2002; 2002WO-JP04166.
 XX PR
 PR 27-APR-2001; 2001JP-0130922.
 XX PA
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX PI
 PI Kubota M, Nishimoto T, Higashiyama T, Watanabe H, Fukuda S;
 XX MI
 MI Miyake T;
 DR WPI; 2003-067722/06.
 DR N-PSDB; ABV75608.

XX PT Production of isomaltose used for manufacturing e.g. beverages, health
PT foods and cosmetics by using specific non-reducing end
PT alpha-1,4-glucosyl linkage-containing saccharide and alpha-isomaltosyl
PT glucosaccharide synthase
XX
XX Disclosure: Page 103-111; 121pp; Japanese.
XX
CC The invention relates to a novel method for the production of isomaltose,
CC comprising treating a saccharide having alpha-1,4-glucosyl linkage at the
CC non-reducing end and a glucose polymerization degree of up to 2 with
CC alpha-isomaltosyl glucosaccharide synthase, treating the product(s) with
CC isomaltose-releasing enzyme and collecting the obtained isomaltose.
CC Isomaltose is useful for manufacturing beverages, health foods, feeding
CC materials, cosmetics, drugs and cooking sauces. The sequence represents a
CC polypeptide used in the invention. No further information about the
CC sequence can be obtained from the specification.
XX
SQ Sequence 1284 AA:
Query Match 3.2%; Score 149; DB 24; Length 1284;
Best Local Similarity 18.8%; Pred. No. 0.028;
Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;
QY 105 TLTYPPSLSPTIYHLDGVEG--GIDFTPHNGTKKIINTVAEYNKLSADSGSIHSH- 161
DB 54 TLTVNGAEPSSDLLIVQVQNGILKVDYRPNISITPSAKTPLMDPNKWTSAVGATINTTA 113
QY 162 -----LTNNALVEI-----HTANGRWVRDIYLPQGPDLGKWKVRFVSAGYSSTVF 207
DB 114 NPMITTSNMKIEITKNPVRMTVKKADGTTL--FWPEPSGGGVFDGVRFLHATGDN--M 168
QY 208 YGDRVTVLSVGNLTLLFKYVNGQWFRSGLENRRITVIAQHINSAELPAHWIVPGLMLVIKQ 267
DB 169 YGIRSF-----NAFDSGGDLLRNSNHAH----- 193
QY 268 GNLGRRLNDIKIGAPGELLHLLTIDIGMLTTPDRDFDAKDEAHREYFQ--TIPVSRMIV 325
DB 194 ---AGEQGD---SGGLTIWSTAGYGLLVSDGYPYTDSTTGQMEFYGGTTPBGRRYA 246
QY 326 NNYAPLHL-----KEVMLPTGELLTMDPGNGGWHSGTNR-----Q 361
DB 247 KONVEYIIMLGPKEIMTDVGE-ITGKPPMLPWSLGFNFNFWDNQTEFTNNVDYTRAK 305
QY 362 RTGKELVSHGIDNANYNSTAGLGENSHYVVAQLAAHNSRGNYANGIQVHG----- 414
DB 306 NIPIDAYAFDYKKYGETNYGEPAWNTNFPASSTSLAKTMD-AGKIMIGITKPRIV 364
QY 415 -GSGGGGIYV---LDSTLGNFSEHVGHNHGLHYVDGFGKGSVHRSAE---NNNSTGWKDG 468
DB 365 TKDASANVTTOGTDATNGCYF--YFCHN---EYQDYFIPVTVRSDIPYNANERAWFNH 418
QY 469 DK---KREIPNYPYPSOTNEKS-----CLNNCOEPPDFGHKFGFDAMAGSP--- 511
DB 419 STDALNKGVGNWDEIDKVSFGGALYFNGTNTTGHMSQTHTEGER-----ATSGARVW 474
QY 512 -----FSAANRF--TMVTP-----NSSAIIQRFFENKAVFDSRSSTGFSKNW 551
DB 475 QTARTYFPCAQRAYATTLWSGDIGIQYKNGERINWAAGMQ---EQRAVMLSSVNNQGVK 531
QY 552 ADT-----QEMEPY-----EHTIDRA 567
DB 532 MDTGFGNOODGTNNPNPDILYARMWOFSAITPFRVHGNHNOQROPNWYFGSTAEEASKEA 591
QY 568 EQI-----TASVNELSESKMAELMAEY---AVVK--VHMNN-CNW----- 601
DB 592 IQRLYSILPYMYAYERSAYENGVLRLQVAYPTDAAVKNTDAMWFGDWLLAAPVDK 651
QY 602 ---TRNIYIPTASDN--RGSILT-----INHEAGVNSYLFIDKGVYSGQYKXSF 648
DB 652 QQTSKDIYLPSPGWDIYARGNAITGGQITRYSVNPTLMDPLFI-----KKGA 700
QY 649 VSDGQFWKERDVVDTREARKPEQFGVPVTVLGV--YYDPEGTLSYIYPAMYGAYGFTYS 706

DB 701 IIPQ--KVQDYVQASVTSVDVDFDTTQSSFTYYDDG-----ASYNYESGTYF 750
QY 707 DDSQNLSDNDCQLQVDTKEGQLRFLRANHRANNTVYKFKHINVTPTSQPTQATLVCKNKI 766
DB 751 KONMTAODNG-----SGSLFTLLGAKSGSYTPALQSYI---VKLHGSAGTSVTNNS- 798
QY 767 LDTKSLTPAPEGLTYTVNGOALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDMIVGOE 836
DB 799 -----AAMTSTASLEALKAAGEG-----NATGKD 823
QY 827 VYVD 830
DB 824 IYGD 827
RESULT 8
ABB47329
ID ABB47329 standard; Protein; 1530 AA.
XX AC ABB47329;
XX DT 05-FEB-2002 (first entry)
XX DE Listeria monocytogenes protein #33.
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX KW vitamin B12; bacterial infection; disease.
XX OS Listeria monocytogenes.
XX PN WO200177335-A2.
XX PD 18-OCT-2001.
XX PF 11-APR-2001; 2001WO-FR01118.
XX PR 11-APR-2000; 2000FR-0004629.
XX PA (INSP) INST PASTEUR.
XX PI Buchrieser C, Frangeul L, Couve E, Rusnok C, Fsihi H, Dehoux P;
PI Dusserget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
XX PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX PT and prevention of Listeria and related bacterial infections, and
XX PT related polypeptides
XX PS Claim 6; SEQ ID No 34; 192pp; French.
XX CC The present invention relates to the genome sequence of Listeria
XX CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX CC it are useful for selecting probes and primers for detecting genes in L.
XX CC monocytogenes and related organisms, and for studying genetic
XX CC polymorphisms and other genomes. The present sequence is a protein
XX CC encoded by the genome sequence of the present invention. Proteins
XX CC expressed from the genome sequence are useful for raising specific
XX CC antibodies. Identification of L. monocytogenes and related organisms,
XX CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX CC B12. The genome sequence and proteins encoded by it are also useful for
XX CC selecting compounds that regulate gene expression and cell replication
XX CC and modulate L. monocytogenes-related diseases. In addition, the genome
XX CC sequence and proteins encoded by it are useful in pharmaceutical and
XX CC vaccine compositions for the treatment or prevention of infections by L.
XX CC monocytogenes and related organisms.

```

Query Match      3.1%; Score 145; DB 23; Length 929;
Best Local Similarity 19.8%; Pred. No. 0.035;
Matches 165; Conservative 107; Mismatches 288; Indels 272; Gaps 49
Qy 176 RWRDYILPQG---PDLEGKMYRFVSSAGYSSTVFYFGDKVTLTSGVNTLLFKYVNGQWR 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 RW-----YLMGGDPLDLRHDYKELTGTPTVPVPKAFG-----LMYSE---FGYON--W-- 255
Qy 233 SGELENNRITYACHTWSAELPAHWTVPGLNLVIKO-GNLSGRINDIKIGAPCELLHTI-- 290

```

XX	09-JAN-2002; 2002WO-JP00052.
PF	
XX	
XX	12-JAN-2001; 2001JP-0005441.
XX	
PA	(HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX	
XX	Kubota M, Maruta K, Yamamoto T, Fukuda S;
DR	WPI; 2002-520129/55.
XX	
PT	New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
PT	the production of cyclic tetrasaccharide gum -
XX	
PS	Disclosure; Page 137-141; 144pp; Japanese.
XX	
CC	The invention describes novel microbial polypeptides having
CC	alpha-isomaltosylglucosaccharide synthase activity. The proteins are
CC	useful for producing a sugar (I) having at least three glucose units
CC	connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
CC	end. The invention also describes a method for the production of the
CC	cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
CC	glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
CC	(I-) (II) by treatment of (I) with alpha-isomaltotransferase.
CC	(II) and similar sugars in crystalline or syrup form are used as
CC	a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
CC	and colour stabiliser for foodstuffs, cosmetics and drug compositions.
CC	This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
CC	synthase.
XX	
SQ	Sequence 965 AA;
	Query Match 3.18; Score 145; DB 23; Length 965;
	Best Local Similarity 19.8%; Pred. No. 0.037;
	Matches 165; Conservative 107; Mismatches 288; Indels 272; Gaps 49;
QY	176 RWRDYLPG--PDLEGKMRVFSAGYSTVFYGDKRVTLSVGNLTLLFKYNGQWFR 232
DB	246 RW----YLMGDDLPDLRHDMETLTPPPPKAFG-----LWSE---FCYDN--W-- 289
QY	233 SGLENNRITYAQHIWAELPAHWIVGLNLVIKO-GNLSGRINDIKIGAPCELLHTD- 290
DB	290 -SEVDNT-----IAGLRADFP-----VDGMLDVOMFGGVTAADDTRMG-----TLD 332
QY	291 -DIGMLTTPDRF--DFADKEAHREYFTQIPVSAMVNMYNAPLHLKEVWLPTGELLTD-- 346
DB	333 WDISRPDPACKIADLAEDVG-----IPIESYYCRNLPEHAR--MAADGYLVRSGC 384
QY	347 -----MDP--GNGG-----WHSTMTQRIOCKELVSHOIDNAN---YGLN 380
DB	385 ATCPPVYLTGNPMWGKGIMDWTQPEAGVMHDQRQLVDGVLHGLWDIGEPEMYDPN 444
QY	381 S-TAGL--GENSH-----PYVVAQLAAHSRGNYANGI 410
DB	445 DWTAGTVPGKHAAHYINAYNLLWASIDAGYONGVKRFMLTRA--AAGI 496
QY	411 QVHG--GSGGGGIVTLDTSLGNEFSHEVGHNY-GLGHY---VDGFKGSVHRSAENNNSTWG 465
DB	497 QRHCAGWSADIGSTMALGSQQNAQMMSNSGDIYGSIDIGGFREM-AODGVNELYTQ 555
QY	466 WDGDKKRFIPNFYPSQTNRKSCLNNQCQEPDPDGHKHFGFDAMAGGSPESAARNFTMYTENS 525
DB	556 WFADSASFDTPLRPHTDNLNCLET-----SPDSIGD---VASNR 592
QY	526 SAILIQFFENKAVPDSRSSGFSKNADTQEMEYEHTI-DRABOI----- 570
DB	593 ENLVRRY-----ELAPFYYSLAHRAHFGEPLAPPLYTYQN 629
QY	571 TASYNLESSEKMAELMAYAVKVVMWNGNTRNIYIPTASADNRGSTILTINHEAGYSY 630
DB	630 DDHVREMGHQH---IGRDLLIAIVAGEGERDYILPA-----GEW 668
QY	631 LFIGNDEKYVSQGIKSFSDGQFWKERDVYDTREARK-----PEQFGVPVTLVGYD 684

Db 669 IDIHTNERIQSTG---QWIDNVPWM--RDGVFTLPAYARAGALIIPRAVDASTKDITGKR 723
 Qy 685 PEGTSLSYIPAMYG---AYGFT-YSDSQNTSDNCQLOVDTEKQGLRFLRANHRANT 740
 Db 724 EDAVRNELIATYVADVASDFLYEDDAGATTAYADGAVRTT---QISQSLTNGVATVT 779
 Qy 741 V--MNFHFHINVPTESOPTOATLVCCNKKILDTKSLTPEAGELTYVNGQALPAKEN----- 793
 Db 780 VGRASGTSGAPS--TRPTVVELVTDGTQASTVSLGVP--LHEHANKAFAFDASSGWYNA 836
 Qy 794 -EGCIV-----SVNSGKRYCLPVGQRSGYSL-----PDMVQGEVYVDSGAKAKVLLSD 841
 Db 837 GGGLVAKAASSSVNTAKTFSTLGEESVWATFSCENATTTFCQSYV----- 884
 Qy 842 WDNLNRYRGEF--VGNVNPADMKKYK-----AWNGQYLDKFSKPSRMVRYK 886
 Db 885 -----VGNVPOLGNMSPADAVKLEPSAYPTWTGVVRNL--PPSSTVEMK 926

RESULT 11
 AAB10667
 ID AAB10667 standard; Protein: 2057 AA.
 AC AAB10667;
 XX 19-JAN-2001 (first entry)
 DT L. mesenteroides alternan sucrose protein.
 XX Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;
 KW syrup.
 XX Leuconostoc mesenteroides.
 OS DE19905069-A1.
 PN 10-AUG-2000.
 PD 08-FEB-1999; 99DE-1005069.
 PF 08-FEB-1999; 99DE-1005069.
 PR (PLAN-) PLANTEC BIOTECHNOLOGIE GMBH.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX Kossmann J, Welsh T, Quanz M, Knuth K;
 DR WPI: 2000-550294/51.
 DR N-PSDB; AAA97904.
 XX New nucleic acid encoding recombinant Leuconostoc mesenteroides
 PT alternan sucrose protein and methods of alternan and fructose
 PT production -
 XX Claim la; Page 30-36; 64pp; German.
 XX This invention describes a novel nucleic acid molecule (I) encoding an
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the
 CC glucosyltransferase group) The recombinant, purified alternan sucrose
 CC gene is useful for the fermentative production of alternan (a
 CC carbohydrate) and/or fructose by secreting the enzyme into a
 CC saccharose-containing culture medium. Alternatively, the enzyme is
 CC contacted with a saccharose-containing solution. The alternan and/or
 CC fructose is then isolated from the medium. Cosmetic products or
 CC foodstuffs containing alternan can be produced. Recombinant production of
 CC alternan sucrose is advantageous as it provides a cost effective means of
 CC producing fructose for high fructose containing syrups, production of
 CC which previously has been achieved by costly production from maize
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan
 CC sucrose protein which is described in the method of the invention.
 XX Sequence 2057 AA;
 SQ

Query Match 3.1%; Score 144.5; DB 21; Length 2057;
 Best Local Similarity 17.6%; Pred. No. 0.14; 346; Indels 485; Gaps 52;
 Matches 203; Conservative 122; Mismatches 122; Indels 485; Gaps 52;
 Qy 22 TAADNNSAIYFNTSQPINDLOGSLAAEYKFAQSILPAHPKEGDSQPHLTSLRKSLL-- 79
 Db 910 TTTQDNSQYKN--QGIGVIVGN-NPDLKLNDDKITLHM-----GKAHKNQLYRALVLSN 962
 Qy 80 ---VRPVXADKTPVQVEARDNNKIL-----GTLTLYP-PSLPTPTIHLDG-- 123
 Db 963 DSGIDVYSDDKAPT-LRTNDNGDLIFHKTNTFKQDGTIINYEMKGSNLALISGYLGW 1021
 Qy 124 VPEGGIDTPHNGTKKIINTVAENVKLSDSGSSISH--LTNNALV----- 169
 Db 1022 VPVGASD-----SDARTVATESSSN-DGSVPHSNAALDSNVIEGFSNQAMTSP 1073
 Qy 170 -----INTANGRVROI-----YLPQGPDLCKMVRVSS-----AGYSSTVFYGD 210
 Db 1074 EQSTNVIAATKANLKFELGITSFELAPOYRSSGDTNYGCM-SFLDSFLNNGYAFDTRY-D 1131
 Qy 211 RKVTLSVGNLTLLFKYVNGQWFRSGELENNRITYAQHIWSAELPAHWI-----VPLGNLV 264
 Db 1132 LGFNKADGNPNPKYGTDDLRNA-----JEAUHKMGQAIAWMVDDQIYALPGKEV 1184
 Qy 265 I-----KQGN-----LSG-----RLNDIKI 279
 Db 1185 TATRYDERGNQLKDTDFYNLLVAVANTKSSGVYQAKYGGEFLDKLREYPSLFKONQVST 1244
 Qy 280 GAP-----GELLHTIDIGMLTTPDRPDFAKOKEAHREYFOTIPVSRM 323
 Db 1245 GQPIDASTKIKOWSAKYMNGTNIH-----RGAYYVLKD-WATNQYFNIAKTREV 1293
 Qy 324 IVNNTAPLHLKEVMLPTGELL-----TMDPGNGGWHSGTMRQRICKE 366
 Db 1294 ----FLPQLQNKDAQTGFISDASGVKYYSISGYQAKDTFIEDGNGNYY----- 1339
 Qy 367 LVSHGIDNANYGLNSTAGIGENSHPYVVAQLAAHNSRGN-----ANGIOVHGSGGGGIVT 423
 Db 1340 ----FDKGYMVRSQG---ENPIRTVETSVNTRNGYFMPNGVELRKFG----- 1384
 Qy 424 LDSTLGNFEFSHEVGHNYGLGHYVDGFKGSVHSANNSNTHGWDGDKKRFIPNFYPSQTN 483
 Db 1385 -----TNSGNVYTFDDGQKMRVKYINDDAN 1411
 Qy 484 EKSCLNNOCEPFDG-----HKFGFDAMAGSPFSAANRFTM---YTPNSSAIIQREFE 534
 Db 1412 NPYHLN-----VDGTMGRGLFKFSDTL---QYFASNGVQIKDSYAKDSKG----- 1454
 Qy 535 NKAVFDS---RSSTGFCKWADTQEMEPYEHYIDRAEQITASVN----- 575
 Db 1455 NKYYFDSATGNNDTG---KAQTDWNGYIITIDSDANNITGVNTDYATYITSSLEDGL 1510
 Qy 576 -----ELSESKMAELMAEYAVVYVVMWNGNTRNIYIPT 609
 Db 1511 PANAPYGVVTKDQNGNDLAKQYINHTKQYEGQOVVTKQYTDTSKGVSN-----LITF 1563
 Qy 610 ASADNRGSILTINHE-----AGNYSYLFINGDEKVVSGYKKSFYSD 651
 Db 1564 AGDLOGQLRWDSRALTWTPTKTNQIISFIYANRNDGLFLNAPYOV--KGYQLAGSN 1621
 Qy 652 -----GQFWERDVVDTRARKPEQ----- 671
 Db 1622 QYKGOQVVIAGVANVSGKDWLSISFNGTQYITDSQALNTNETHDMNQKVFVNTTSLNDGL 1681
 Qy 672 -FGVP-----VTLVGYVDPSTGLSSYIYPAMYGAYGFTYSDSQN 711
 Db 1682 FLNAPTRQPGYKLAGLAKYNNQTVYVQYFDDGTVVGLGQVTVWDNHAQMO 1741
 Qy 712 LSDNDCQLQVDTKEGQLRFLANHRANNTVMKHFHINVPTESQPTOATLVCNKKILTDS 771
 Db 1742 VSDTDOQLVYN-----SNCRNDGLFLNAPYRQGSQLI----- 1774

24-APR-2001 (first entry)
M. catarrhalis les1 200kDa protein SEQ ID NO:11.
Moraxella catarrhalis strain Q8; major outer membrane protein;
200kDa outer membrane protein; antibacterial; immunogenic; infection;
otitis media; detection.
Moraxella catarrhalis.
WO200107619-A1.
01-FEB-2001.
26-JUL-2000; 2000MO-CA00870.
27-JUL-1999; 99US-0361619.
(CONN-) CONNAUGHT LAB LTD.
Loosmore SM, Sasaki K, Yang Y, Klein MH;
WPI; 2001-159722/16.
N-PSDB; AAF59105.
New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
useful in protective vaccines and for diagnosis -
Claim 1; Fig 5A-Y; 247pp; English.
The present invention describes an isolated and purified nucleic acid (I)
that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
The 200 kDa outer membrane protein (II) has antibacterial activity and
can be used in vaccines. (II), and its truncated versions, are used as
immunogenic compositions and vaccines to protect against M. catarrhalis
infections, particularly otitis media in humans. (II) is also used as
antigen in immunoassays for detecting specific antibodies (Ab), and its
generate Ab. (I) are used for recombinant production of (II) and its
fragments are used as probes for identifying/cloning 200 kDa protein
genes from other strains, and for diagnostic detection of M. catarrhalis.
(I) makes possible production of large amount of recombinant immunogens.
Expression of truncated versions of (II) reduces toxicity of the protein
towards the Escherichia coli host. The present sequence represents the
M. catarrhalis les1 200kDa protein, which is given in the exemplification
of the present invention.

Query Match 3.1%; Score 144; DB 22; Length 2314;
Best Local Similarity 20.1%; Pred. No. 0.18;
Matches 212; Conservative 129; Mismatches 388; Indels 326; Gaps 52;

QY 9 TILAPLAIGVFSA-ADNNSAIVNTSOPINDLQGSIAAEVKEFAQSOILPAHPKE 63
DB 47 TRVATLAILVIGATLINGSAYAQNNSKIAFGTGG--NNDNASASNEASIAIGSLAKARANQ 104
QY 64 ----GDSOP-----HLTSLRKSLLVPRVKADKTPVQVEARD---DNKKIL 103
DB 105 ATAIGSRPDRPQAQNAQKAGSHAGK-RESIAIGSDVLAEGDASIAIGSDDLDRN--- 160
QY 104 GILTYLPPSLPDTYHLDGVPPEGIDTPPHNGTKKIINTVAENVKLSADSGSSHS--H 161
DB 161 STNSKYPNGULLSTLQNHVLRQ-----IPDSNGSOKYRRTAAE-GHASTAVGAMAYAKGH 215
QY 162 LTNALVEIHTANGVRVDIYLPQGDLEGKRVFSSA--GYSTFYVGDVKVTLVSG 218
DB 216 FA-NAPGTRSTAEAGNYSIAVGUTAKAE-KGYTIAIGSNAQAINYGALGADTRVDLDYG 273
QY 219 NTLIF-----KYV---NGWFRSGELENRRITYAQHIWISAEPLPAHWIPLG 261
DB 274 IALGYSQLINNNNNNNNNKAYPEGNGSNIKSKATGNCLF---SIGSTIKRKLIINVA 330
QY 262 -----NLVIFOGNLSGRNDI-----KIG-----APGELLHTID 291

DB 331 GYEDTDVAVNAQLKAVENLAKRQITFRGDDNGTVGKKKLGETLTKGGETQADKLNNN 390
QY 292 IGLMTPRD--RFDFAXDKAEAREYFQIPVSRMIV-----NNYAPHLREY-MLP 335
DB 391 IGVVTDNNTGLKVLAKNLG---LETVSTKNLTASEKVTGSGNNTAELOSGLTFTP 446
QY 340 TCELLTD-----MDPONGCWBHSGTMRQIRGLKELVSHGIDNANYGLNSTAGLGEN 388
DB 447 TTNASTDKTVYGTGDKFTDNTALEDTT---RITDKI-----GFSNKAQTVDE 494
QY 389 SHPYV-VAQLAAHNSRGNVANGIQVHGSGGIVTLTDLSTLGNFSEHVEGNYGLGHVVD 447
DB 495 NKPYLDKDKLVGNSTLN-----NGG---LTVNNTIG-----GSKKQIQVAD 534
QY 448 GFKGSVHRSANNNSTWGN-----DG-----DKRF----- 473
DB 535 GIKFADVNVVNSAAKFGTTRITEEIGFADADGVKDKSPYLDKKQLQVGVKITKDSG 594
QY 474 -----IPNFYPSOTNEKSLN---NOCDEPPDGHKFGPDAM-AGGSPFSAANRTM- 520
DB 595 INAGDOKISN-VKDATDDTDAVYKQLQVQODADGALQSIRDEKQETISNLSNG 653
QY 521 YTPNSAIIQREFENKAVFDSRSSTGFSKWNADTOEMEPYEHTIDRABQITASVNELS 580
DB 654 NTPNTFETITFAGENGISISNDIAKGVKVGID-----PINGLTTP 694
QY 581 KM---AELMAEYAVVVMWNGNWTNRI---YIPT-ASADNRGSILT----- 620
DB 695 KLTVGSDKDKTQLVIEQVAGSNDTKNIIRGLSPTLPSITNAGGVRTTEQGNITSDBK 754
QY 621 -----INHEAGVNSYLFINGDEKVVVSQGY-----KKSFVSDQGF 654
DB 755 SKAASIGDILNTGFNLKNNNSVGFVSITYNTVDFIDGNATTAKVYDETNQTSKYTYDVN 814
QY 655 WBERVDVDTREARKPKQFPQVPTLVGYDPPGTLSSYIYPAMYGAYFTYSDSDSLS 714
DB 815 VDEKTIETLGDNGKTKNKGKVTTLT-----TTNA-----NGKATNFST 853
QY 715 NDCQLQVDYTKEGRLPRLANHRANNTVMKHNIVPTESQPTQATLVCMNKLDTKSLTP 774
DB 854 TNDALVNAKD-----IAENL--NTLAKIHTTGTADTALQTFVKKDKGATDDEITV 905
QY 775 APBGLTYTVNGQ---ALPAKRENGCIVSVNSGKRYCLPVGQRSGYSLPDWIVYQEVYDS 831
DB 906 GKDG---TQNGKTVNTLKLKGENGLVATNK-----DGTVTGINTQS 945
QY 832 GAKAKVLLSDWMLSYNRIGFVGVNVPADMKVK 866
DB 946 GLKA-----GDSITLTKDG--LSINKPASNEQIQ 972
RESULT 14
ABG30538
ID ABG30538 standard; Protein; 1251 AA.
XX
AC ABG30538;
XX
DT 07-OCT-2002 (first entry)
XX
DE Alpha-isomaltosylglucosaccharide synthase #2 mature protein.
KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
KW cosmetic; drug composition; enzyme.
XX Unidentified.
OS
XX
PN WO200255708-A1.
XX
PD 18-JUL-2002.

[illegible]

[illegible]

Search completed: October 8, 2003, 16:24:44
Job time : 59.9419 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: October 8, 2003, 16:19:57 ; Search time 47.5173 Seconds
(without alignments)
4811.606 Million cell updates/sec

Title: US-10-002-309b-2
Perfect score: 4709
Sequence: 1 MKLYLSTILAPLAIGVFS.....ANNGQYLDSPSRMRVYK 886

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4709	100.0	898	082882	082882 escherichia
2	1548	32.9	310	092AL1	Q92all escherichia
3	686.5	14.6	1002	09AGX6	Q9AGX6 vibrio chol
4	671	14.2	1335	16 Q9KN18	Q9KN18 vibrio chol
5	571	12.1	875	16 Q8D5P4	Q8D5P4 vibrio vuln
6	290	6.2	957	16 Q9KTH2	Q9KTH2 vibrio chol
7	166.5	3.5	1217	16 Q97GW3	Q97GW3 clostridium
8	161	3.4	1038	2 Q45780	Q45780 bacteroides
9	159.5	3.4	2367	2 Q46034	Q46034 clostridium
10	158.5	3.4	2204	5 Q81L70	Q81L70 plasmodium
11	156.5	3.3	1157	16 Q97PF2	Q97PF2 clostridium
12	154.5	3.3	2367	2 Q9F931	Q9F931 clostridium
13	150	3.2	3692	2 Q8KRR3	Q8KRR3 fusobacteri
14	149	3.2	1284	2 Q8RQU9	Q8RQU9 bacillus gl
15	148.5	3.2	2468	16 Q912M3	Q912M3 pseudomonas
16	148	3.1	1021	2 Q46085	Q46085 clostridium

17	148	3.1	2937	16 Q8D990	Q8D990 vibrio vuln
18	147.5	3.1	1530	16 Q8Y479	Q8Y479 listeria mo
19	147.5	3.1	2306	16 Q8F107	Q8F107 leptospira
20	147	3.1	1093	16 Q8EV69	Q8EV69 mycoplasma
21	147	3.1	1377	16 Q98Q28	Q98Q28 mycoplasma
22	147	3.1	1386	16 Q92DL0	Q92DL0 listeria in
23	146.5	3.1	1196	17 Q8TS48	Q8TS48 methanosarc
24	146.5	3.1	2057	2 Q9RE05	Q9RE05 leuconostoc
25	145.5	3.1	2189	5 Q8IRV6	Q8IRV6 plasmodium
26	143.5	3.0	1395	2 Q9AISO	Q9AISO staphylococ
27	143	3.0	1284	16 Q8EV71	Q8EV71 mycoplasma
28	140	3.0	1135	1 Q9P9D1	Q9P9D1 uncultured
29	139	3.0	1110	17 Q8PXT0	Q8PXT0 methanosarc
30	138.5	2.9	1270	16 Q8XAN9	Q8XAN9 escherichia
31	138.5	2.9	1575	2 Q9LCH3	Q9LCH3 streptococ
32	138	2.9	958	17 Q8TL14	Q8TL14 methanosarc
33	138	2.9	1964	2 Q8KQK9	Q8KQK9 moraxella c
34	138	2.9	2675	5 Q813F9	Q813F9 plasmodium
35	138	2.9	2894	17 Q58791	Q58791 methanococ
36	137.5	2.9	2529	16 Q25579	Q25579 helicobacte
37	137	2.9	1480	2 Q8GC13	Q8GC13 lactobacill
38	137	2.9	5559	16 Q82KG6	Q82KG6 salmonella
39	136.5	2.9	1457	2 Q9F289	Q9F289 versinia pe
40	136.5	2.9	3705	2 Q9F285	Q9F285 versinia pe
41	136.5	2.9	3705	16 Q8ZHA1	Q8ZHA1 versinia pe
42	136.5	2.9	3710	16 Q8CZU2	Q8CZU2 versinia pe
43	136	2.9	1677	16 Q8RGN7	Q8RGN7 fusobacteri
44	136	2.9	1980	3 Q05747	Q05747 saccharomyc
45	136	2.9	2167	16 Q92EK5	Q92EK5 listeria in

ALIGNMENTS

RESULT 1
082882 ID 082882 PRELIMINARY; PRT; 898 AA.
AC 082882: Q9ZGUL;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Plasmid POL157 DNA, complete sequence.
GN L7031.
OS Escherichia coli O157:H7.
OC Plasmid POL157.
OC Bacteria, Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7;
RX MEDLINE=98290540; PubMed=9628576;
RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA Yutsudo H.C., Kubota Y., Yamauchi Y., Iida T., Yamamoto K., Honda T.,
RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
RA Shinagawa H.;
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
RT outbreak";
RL DNA Res. 5:1-9(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RX MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
RT of Escherichia coli O157:H7";
RL Nucleic Acids Res. 26:4196-4204(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RA Brunder W.;

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB011549; BAA31757.3; -;
 DR EMBL; AF074613; AAC70099.1; -;
 DR EMBL; Y11831; CAA72517.1; -;
 KW Hypothetical protein; plasmid.
 SQ SEQUENCE 898 AA; 99548 MW; 3C1AE23E3EAE1FAB CRC64;

Query Match 100.0%; Score 4709; DB 2; Length 898;
 Best Local Similarity 100.0%; Pred. No. 1.4e-300;
 Matches 886; Conservative g; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 60
 DB 13 MKKYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 72
 QY 61 PKGDSQPHLSLRKSLLLVRPVKADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 120
 DB 73 PKGDSQPHLSLRKSLLLVRPVKADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 132
 QY 121 LDGVEGGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 180
 DB 133 LDGVEGGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 192
 QY 181 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKVTSYVNTSOPINDLOGSLAAEVKFAQSILPAH 240
 DB 193 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKVTSYVNTSOPINDLOGSLAAEVKFAQSILPAH 252
 QY 241 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKVTSYVNTSOPINDLOGSLAAEVKFAQSILPAH 300
 DB 253 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKVTSYVNTSOPINDLOGSLAAEVKFAQSILPAH 312
 QY 301 RFDFAKDEAHREYFOTIPVSRMIVNNYAPLHLKEVMLPTGELLTMDPGNGGHSCTMR 360
 DB 313 RFDFAKDEAHREYFOTIPVSRMIVNNYAPLHLKEVMLPTGELLTMDPGNGGHSCTMR 372
 QY 361 QRIKELVSHGIDNANYGLNNTAGLGENSHYPVYVAQAAHNSRNYANGIQVHGSGGGG 420
 DB 373 QRIKELVSHGIDNANYGLNNTAGLGENSHYPVYVAQAAHNSRNYANGIQVHGSGGGG 432
 QY 421 IYVLDSTLGNFSEHVGNYGLGHYVDCFGKSVHRSANNNSTWGDGKKRFPNFPYS 480
 DB 433 IYVLDSTLGNFSEHVGNYGLGHYVDCFGKSVHRSANNNSTWGDGKKRFPNFPYS 492
 QY 481 QTNKESCLNNOCQEPFDGKFGFDAMAGGSPFSAANRFTMTTPNSSAIIOFFENKAVFD 540
 DB 493 QTNKESCLNNOCQEPFDGKFGFDAMAGGSPFSAANRFTMTTPNSSAIIOFFENKAVFD 552
 QY 541 SRSTGFSKWNADTOEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVVYVHMNGN 600
 DB 553 SRSTGFSKWNADTOEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVVYVHMNGN 612
 QY 601 WTRNIYIPTASADNRGSLITINHEAGYNSYLFINGDEKVVSGQYKKSFPVSDGQFWKERDV 660
 DB 613 WTRNIYIPTASADNRGSLITINHEAGYNSYLFINGDEKVVSGQYKKSFPVSDGQFWKERDV 672
 QY 661 VDTREARPEQGVPTTLVGVYDPEGTLSSVYIPAMYGAYFTYSDSDNSLNDQCOLQ 720
 DB 673 VDTREARPEQGVPTTLVGVYDPEGTLSSVYIPAMYGAYFTYSDSDNSLNDQCOLQ 732
 QY 721 VDTKEQLFRLANHRANTVANKFHINVPESQPTQATLVCKNKLDTKSLTAPGLT 780
 DB 733 VDTKEQLFRLANHRANTVANKFHINVPESQPTQATLVCKNKLDTKSLTAPGLT 792
 QY 781 YTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDMIVGQYVYVDSGAKAVLLS 840
 DB 793 YTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDMIVGQYVYVDSGAKAVLLS 852
 QY 841 DWDNLNRYIGETVGVNVPADMKKVKANNGQYLDPSKPSRMYVTK 886
 DB 853 DWDNLNRYIGETVGVNVPADMKKVKANNGQYLDPSKPSRMYVTK 898

RESULT 2

Q92AL1 PRELIMINARY; PRT: 310 AA.
 ID Q92AL1
 AC Q92AL1
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 34.0 kba protein (Fragment).
 OS Escherichia coli.
 OG Plasmid p0157.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EDJ933;
 RA Brunder W.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y11275; CAA72142.1; -;
 KW Hypothetical protein; Plasmid.
 FT NON_TER 310 310
 SQ SEQUENCE 310 AA; 34032 MW; 2AA72C166F315BA6 CRC64;
 Query Match 32.9%; Score 1548; DB 2; Length 310;
 Best Local Similarity 100.0%; Pred. No. 1.2e-93;
 Matches 298; Conservative g; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKKYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 60
 DB 13 MKKYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 72
 QY 61 PKGDSQPHLSLRKSLLLVRPVKADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 120
 DB 73 PKGDSQPHLSLRKSLLLVRPVKADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 132
 QY 121 LDGVEGGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 180
 DB 133 LDGVEGGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 192
 QY 181 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGLENNR 240
 DB 193 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGLENNR 252
 QY 241 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGLENNR 298
 DB 253 IYLPQGPDLGKMYRFVSSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGLENNR 310
 RESULT 3
 Q9AGX6 PRELIMINARY; PRT: 1002 AA.
 ID Q9AGX6
 AC Q9AGX6
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE ToxR-activated gene A protein.
 GN TAGA
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=395;
 RA MEDLINE=21117025; PubMed=11179381;
 RA Karalis D.K.R., Lan R., Keper J.B., Reeves P.R.;
 RT *Comparison of Vibrio cholerae Pathogenicity Islands in Sixth and
 RT Seventh Pandemic Strains.*
 RL Infect. Immun. 69:1947-1952(2001).
 CC -!- SIMILARITY: CONTAINS 1 FIBROTECTIN TYPE III-LIKE DOMAIN.
 DR EMBL; AF125733; AAK20748.1; -;
 DR Interpro; IPR003961; FN_III.
 DR Pfam; PF00041; fn3: 1.
 DR SMART; SM00060; FN3; 1.

KW Hydrolase. 1002 AA; 114645 MW; 6A239DFB6408407F CRC64;
SQ SEQUENCE 1002 AA; 114645 MW; 6A239DFB6408407F CRC64;
Query Match 14.6%; Score 686.5; DB 2; Length 1002;
Best Local Similarity 24.0%; Pred. No. 2.8e-36;
Matches 230; Conservative 157; Mismatches 275; Indels 295; Gaps 40;
QY 27 NSAIYFNTSOPINDLOGSLAAEVKFAQSOILPAHPKEDGSOPLHLSLAKSLLLVRP-VKA 85
DB 137 NEIDYVHTTEIKNVYSGLEGEVTVQTHVI--SPEGRKNEPEITGRDALILFKPSIKN 194
QY 86 DDKTPQVQEARDDNKKILGTLTLYPPSLPDTIYHLDGVPEGGIDFTPHNCTKKIINTVA 145
DB 195 SSSILMKIYSEOG--LTSKVYMKSPMLPKTDQPID-----ID----- 230
QY 146 EYVKLSASGSSIHSHLTNNALVEIHTANGRWVRDIYLPQGDLEGKMYRFVSSAGYSST 205
DB 231 ----- 230
QY 206 VFYGDVKVTLVGNLTLLFKYVNGWFRSGELENRR-ITYAQHINSAELPAHWIVPGLNIV 264
DB 231 -----ENNVKVSISNSYSAEIPAKWKMKGMSLH 259
QY 265 IKQGNLSGRL-----NDIKIGAPGELLHTIDIGMLTTPDRDFDPKDKAHR-EYFQTI 318
DB 260 FEDEN--GNLGIIESERIKFSAPSELLIONIDGLMYKPRNIVIKELERTAVDYFOKV 317
QY 319 PYSRMTVNNYAPLHLKEVMLPTGELLTMDPONGGWHSGTHQRIGKELVSHGIDNANYG 378
DB 318 PYSKLIFSDTPIHFKEKITPLNGSVYVTEKSDADIGGWHQGDREAVGKALVSTGINNANLG 377
QY 379 LNSTAGLGENSEHPYVVAQAAHNSRGNANGVQVHGSGGGGIVTLDSTLGNFESHEVGH 438
DB 378 IYASSGYSQ-QYNRLTNHTAHTNIGYVNGVYVHGSGGGGIVTLENTLNENSHELGH 436
QY 439 NYGLGHVYDGFSGVSHRSARENNSNTGWGDKRFLIPNEYPSOTNEKSL--NNOCQEPF 496
DB 437 NYGLGHVYVAC--GTSH---GPDTSWGDYTKRETIANFDMKRSPQSNIRPDNQEVKPF 490
QY 497 -DGHKGFEDAMAGSGFF--SAANRFMTYTPNSAIIQRFENKAVDFSRSSGFSKWNAD 553
DB 491 MDKYTLWDMAMSGYDHQNGIISRYTLHPHYVARIIQDLKNGAVV---INNDYVWD-E 546
QY 554 TQMEPEYHTIDRAEQITASVELSESKMAELMAEYAVYVHVMNGNMTNRIYIPTASAD 613
DB 547 LKNIYVYKGT-----NEKVPKKGVPVVTILGVYDPDKIN-----PSQLYPTYS-- 591
QY 614 NRGSL-----TINHEAGYNSY---LFINGDERKYSQ----- 642
DB 592 NYGNIFDLEKPRSESLKGWQYVYKOVNYLDNRVNTHHMTLNVNRKEEIKRFSYLSPKGKK 651
QY 643 ---GYKK-----SFVSDQOFWKERDQVDTREARKPEQVPTTLVG-----Y 682
DB 652 FEFGLVEDIENKICTGSRSHIYLEDG---KKNP1-----ESKYNDYF---LLSIDGGEIS 701
QY 683 YDPEGL-----SSYIYAMYGA-----YGFYVSDSQNLS--- 713
DB 702 YVPDSTIGESKICSLKMSGTVYGAGFIKGNCSRQIDGVFMNGFQWAFITLQSGVNSTYTW 761
QY 714 DNDQLOVDYTKGQLR-FRLANHRANTVYNNKPHINVPTESTOPQATLVCKNKLDTKSL 772
DB 762 SNECVLKIKDDNNIESISIPNTRIEKNSKNLHLNISREKPIIDINIVYCGSHELTISKV 821
QY 773 TPAPEGLTYTVNGOALPAKENEICYSVNSGKRYCLPVGQSGYSLPDWIVGQYVYDVG 832
DB 822 SDNPD-----IKLLKGP1I-----VQGEHCYT-----SYEPKLPSP 852
QY 833 AKAKVLSDMDNLS-YNRIGFVGNVNPADMKVKKAWNGOYL---DFS-KPRSMRVV 884
DB 853 -----WFKHYDNFEPKNEINHELKGRVND-----NDEYICRFNFSDSREMKFV 897

RESULT 4

O9KN18
ID O9KN18 PRELIMINARY; PRT: 1335 AA.
AC O9KN18;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Taga-related protein.
GN VCA0148.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL: AE004356; AAF96061.1; -;
DR TIGR: VCA0148; -;
KW Complete proteome.
SQ SEQUENCE 1335 AA; 149521 MW; 0B339F0DBC2DEFD CRC64;
Query Match 14.2%; Score 671; DB 16; Length 1335;
Best Local Similarity 25.5%; Pred. No. 4.6e-35;
Matches 231; Conservative 108; Mismatches 248; Indels 318; Gaps 36;
QY 39 NDLOGSLAAEVKFAQSOILPAHPK--EGDSQPHLTSL--RKSLLVVPYKADKTPVQVE 94
DB 353 NHLOGSLEGLSITQHTSVSA-PKGNELTGQCHLDAIMNREALLLFTPOQGEINOVRAE 411
QY 95 ARDDNNKILGTLYPPSLPDTIYHLDGVPEGGIDFTPHNCTKKIINTVAEYKNLSLAS 154
DB 412 VFIDGELVQVITLML-PPSAL-----AASQDPENGRMKV----- 443
QY 155 GSSTHSLTNALVEIHTANGRWVRDIYLPQGDLEGKMYRFVSSAGYSTVFYGDRTV 214
DB 444 ---VFESHLA----- 449
QY 215 LSVGNLTLLFKYVNGWFRSGELENRRITYAQHINSAELPAHWIVPGLNIV-----KQGN 269
DB 450 -----WSLPLOWMWMKPGLSURLTDNLGREGV 476
QY 270 LSGRLNDIKIGAPGELLHTIDIGMLTTPDRDFDFAKD-KEAHREYFQTPVSRMTVNNY 328
DB 477 LS--QGEIQFGGAPELVIONIDIGMLMPDRDNTMIONLPTLAADYFQKIPASKLYWADY 534
QY 329 APLHLKEVMLPTGELLTMDPONGGWHSGTHQRIGKELVSHGIDNANTGLNSTAGLGEN 388
DB 535 TPAHFPVYVTPNGVYVYTKSGASTGSHSGMDREAIKAKVYSTGINNANVGVSSAGYSQ 594
QY 389 -----SHPYVVAQAAHNSRGNVAN-----GLOVHGSGGGGIVTLDSTLGNFESHEVG 437
DB 595 YNRRFNH-----ITAHTNVGIYTKKTDLPQVYVHGSGGGGIVTLEATGNWSEHLG 648
QY 438 HNYGLGHVYDGFSGVSHRSARENNSNTGWGDKRFLIPNEF-----YPSQTEKSKLNQ 491
DB 649 HNYGLGHV--PYMASIH---DLESWGWDFAHQRFIGNLHMKGDVYVYTOQGD-----D 696
QY 492 CQEPF-DGHKGFEDAMAGSGP--FSAANRFMTVTPNSAIIQRFENKAVDFSRSTGFS 548
DB 697 IVPPEKDAFRLDQAQNGEQEYVGTISRFTLEHPAQSRKAQRWMMNGFNGLSHSPSGYV 756
QY 549 KWNADTQMEPEYHTIDRAEQI-----TASVNSELS 580
DB 549 KWNADTQMEPEYHTIDRAEQI-----TASVNSELS 580

```
Db 757 QNDQETQRYKAVETDTPKPOOVGVVTVLLGIYDPQNEPNSQIYPLVSYNGVTFELPQP 816
QY 581 KHAELMAEYAVVYVHMNGNNTRNIYIPTASADNRGSILTIINHEAGYSY--LFINGDEK 638
Db 817 EOGEOYLE-----GW-----QAAGD-----LT-QAEIOYNOWQTLIDCQQL 852
QY 639 VV-----SOGYKKSFSVSDQFQKRDVVD--TREARKPEQFGVPTVTLVGYDPEGTLS 690
Db 853 PICRFDYNTNGQATFY--GSLNAQRNVCEGSDMR-----WYNDYQIDSPVG 899
QY 691 SYIYPMATYAGFYTSDDSQ-----NLSND----- 716
Db 900 QYELLSQFGAGNVTYTPNAEIGEYQVLTCLKPHNNHSHDGAGFVRNGRCQEOGVKNNAE 959
QY 717 -----COLQVDTKEGOLREFLANHRANNTVMKFKHINYPTE-- 752
Db 960 GRVWYATRNSEVLSRTLASORCELVEYHRNGSTHALDGNRHKSTESNKRFRVNLMSMEK 1019
QY 753 SQPTQATLVCKNKILDTKSLTPAPEGLTYVNGQALPAKENEGCIVSVNSGKRYCLPVG 812
Db 1020 GVPTQVSLSCSD-LNGTSTLT-----RFTPD-QNPPLDKLKGPII-----IGQ 1060
QY 813 RSGYS 817
Db 1061 EYGYS 1065

RESULT 5
Q8D5P4 ID Q8D5P4 PRELIMINARY: PRT: 875 AA.
AC Q8D5P4;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN W20866..
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT *Complete genome sequence of Vibrio vulnificus CMCP6.*;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF016811; AAC07787.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 875 AA; 97611 MW; ABE604A7979BC96 CRC64;

Query Match 12.1%; Score 571; DB 16; Length 875;
Best Local Similarity 27.3%; Pred. No. 8.8e-29;
Matches 182; Conservative 99; Mismatches 215; Indels 170; Gaps 27;

QY 259 PGLNLVI-----KGNLSGRUNDIKIGAPGELLHTIDIGMLTTPRODFPAKO-KEAHR 312
Db 3 PGLSLRLDNLGRGVLIS--QGEIIFGAGAPELVQNDMGMLIEPRNNTMIQMLPTLAA 60
QY 313 EYFQIPVSRMIVNRYAPLHKVEMLPTEGELLTDMDPGNGGWHSGTRIGRKELVSHGI 372
Db 61 DYFQIPASKLVNADYTPAHFPIVTPMGVYVTDKASGTGGWHSGDMREAIQKAMVSTGI 120
QY 373 DNANYCLNSTAGLGEN-----SHPTVYVAQLAAHNRGNAN-----GIQVHGSGGGGI 421
Db 121 NNANVGILTSAGYSQOYNRRFNH-----ITAHTNVGYITTKDITDLPQIVYHGGSGGGGI 174
QY 422 VTLDSTLGNFEFSHEVGHNYGLGHVYDGFSGVHRSAAENNNSTWGDGDKKRFIPNF---- 477
Db 175 VTLEATTGNWSEHGLGNHGLGHW--PYMASIH-----DWESGWDADPHQRFIGNLHWK 228
QY 478 --YPSOTNKSCLNNOCQEPF--DGHKFGFDAMAGGSP--FSAANRFTMYTPNSSAIORE 532
Db 13 LSLSSNVQVQATTNEABGCIISRLNGEKYCLKRYGERSYSLPSWIYAHVPDQVPSGVSM 72

Db 229 DVYTOQGD-----DIVPPKDAFRFLYDAQNGGEQYVYGTISRFTLEHPAQSRKAQRW 282
QY 533 FENKAVFDSRSTGFSKNADTQEMEPYHTIDRAEQITASYNEL-----SESKMA 583
Db 283 MNNGFNLDSSSPSGYVOWDQAARQYQTVTETDPKPOOTGVAVMTLLGIYDPYENPQOI- 341
QY 584 ELMAEYAVVYVHMN-----GNWTRNIYIPTA-----SADNRGSILTIINHEAGY 627
Db 342 -----YPLVSYNGVTFELPQGVQAGFQPEGVQVADLTFAELSDSWOTLRMDGEOQRV 396
QY 628 NSYLF--INGDEKVSQGYKKSFSVSDQFQKRDVVDTREARKPEQFGV--PVTTLVGY 683
Db 397 CRFTFOAAGDSAVFVGVDQS-----TDRCSGROLQWHINSNMTSAQGDY 443
QY 684 DEPGTLLSYVIYPM-----YCAVGFTYSDSONLS----- 713
Db 444 E-----LLSYGRGAVTYTTPPEVGEVTLCTLNKSTDHGAGFVVGNNCEQISGVNKHKG 500
QY 714 -----DNDQCLQVDTKEG--QLRFLANHRANNTVMKFKHINYPTE 752
Db 501 TWRYALRGDEVLRPSYQYQOGOLDVEFANGCASERVVLNASRHSVND--NKPVNLA 558
QY 753 S-OPTQATLVCKNKILDTKSLTPAPEGLTYVNGQALPAKENEGCIVSVNSGKRYCLPVG 811
Db 559 NGVPTQVRLSCSDRNGETELTRFTPE-----QNPPIADLKGPII-----IG 599
QY 812 RSGYS 817
Db 600 EYGYS 605

RESULT 6
Q9KTH2 ID Q9KTH2 PRELIMINARY: PRT: 957 AA.
AC Q9KTH2;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hemolysin-related protein.
GN VCQ930.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT *DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.*;
RL Nature 406:477-483(2000).
DR EMBL: AE004176; AAF4092.1;
DR TIGR: VC0930;
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND;
KW Complete proteome.
SQ SEQUENCE 957 AA; 103989 MW; 4C6B40FBC755CDAD CRC64;

Query Match 6.2%; Score 290; DB 16; Length 957;
Best Local Similarity 53.8%; Pred. No. 2.9e-10;
Matches 57; Conservative 14; Mismatches 35; Indels 0; Gaps 0;

QY 779 LTYVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDWIVGQYVYDQSAKAKVL 838
Db 13 LSLSSNVQVQATTNEABGCIISRLNGEKYCLKRYGERSYSLPSWIYAHVPDQVPSGVSM 72
```


QY 39 NDLOGSLAEVFAQSILPAHPKEDGSOPLHLSLRKSLILLVR-----PVKADDKTP---V 91
 Db 131 NDMTGSVMA-----IKPDELSG-----ITTAQDMGLSKGIAGSVSVISNDPGGGA 177
 QY 92 QVEARDNNKILGTLTLYPPSLPOTIYHLCQVPEGGIDFTPHNGTKKIIINTVAAEVNK--- 149
 Db 178 QIRIRGSSU-----NASNOLVIDGL---AID---NEGIKGMANGLSMVNPD 221
 QY 150 -----LSDAGSSSHLTHNALVEIHTANGRWTRDYLPQGPDEKGVRFVSSAGYS 203
 Db 222 IETLTVLKASATAIYGSASNGVITITTKKG-----NGQAPSVTYNGSVFS 270
 QY 204 STVFYGRKVTLSVGNLTFLKVVNGQWPSGLENRRITYAQHWSAELPA-----HW 256
 Db 271 KT-----QKRYDVLSDGEYRA-----YANQMGDKLPADLGTANTDW 307
 QY 257 -----IVPGLNLVIRG--NLGRANDIKIGAPGELLHTIDIGMLTTPRDRFPAK 306
 Db 308 QDQIFRTAVSTDRHVSINGFKNLPRVS-----LGYTDONGIVKTSNER-RETA 356
 QY 307 DREAHEYEQ---TIPVSRMIV---NNYAPLHKE---VMLPTGELLTMDMP---GNGG 353
 Db 357 SVNLAFFEDHLKFNINAKFMGNKRYADSCRWRAIDAIDTPRVSNEPDYFTG 416
 QY 354 WHSGTMRQIRIGELVSHGIDNANYGLNSTAGLGE-----NSHPYVVAOLAAHNSRG 404
 Db 417 WQN-----INSTGFSNPDWKYTSNPNPQNPONPLAALCLKNDKG 454
 QY 405 NYANGIQVHGGSGGGIVTLTDLSTGLNEFSEHVECHNYGL-----CHYVDGPKG 451
 Db 455 N-----SNDFCYGVNDVYKPHFLPDLRLHASIGEYAEQT 491
 QY 452 SVHRSANNENSTWGDG---KKREIPNFPYSQTNKESCLNQCQEPDGKHGFGDAMA 507
 Db 492 TIVSPYSFGNYYGNGDVTQYKYNLSYNYQYI--KSLGAND-----FDIMV 538
 QY 508 GGSPPSAANRFTWTPNSSAIIQRFENKAVPDSRSSGTFSK---WNADQOMEPYEHTI 564
 Db 539 GGE------QHEFRN-----CFEGGQGWDSYTOE---PHDA 566
 QY 565 DRAEQITASVNELSESKMA---ELMAEYAVKVHVMNGNWTNRIYIPTASADNRGSLT 620
 Db 567 KLRQATAYATRNLTLSYFQRLYSLNLTFTFTWRDGS-----SRFSKDNRM--- 615
 QY 621 INHEAGYNSYLEFNGDEKVSOGYKSFYSQGFQWKERDQVDTREARKPEQGVPTTLV 680
 Db 616 -----GTFPSLALGWK---IKEENFLKOVNVLSDLKL---LWGIIQCO 654
 QY 681 GYDPECTLSSYIYPAMYAGCF---TYSDDS--QNLSDNDCQLQVDTKEGQLRFLANHR 736
 Db 655 NIGDDPAYLPLXYVNNYEAQYFGDTYTSRPAFNEKLKWEKTTWNAGLDGFLNGR 714
 QY 737 ANNTVNKFERINPTESQPTQATLVNCKNKILDTKSLTPAPEGLTY-----TVNGQ 786
 Db 715 ITGGIDGYF-----RKTMTCVTALRSPMNLQCPDDTEYRFTCKLRYGFSINAK 763
 QY 787 ALPAKE-----NKGCIVSVNSG--KRYCLPVG----- 811
 Db 764 PIVTKDFTWDSYNITWNHNEITKLTTGGDSYIYVEAGDKISKGNNTKVOAHKVYAANS 823
 QY 812 -----QRSYSLPWIHVCQEVYVDSGAKAKYLLSDWNLSVNRIGETFGVGNV 858
 Db 824 FYVSRGNNTKVOAHKVYAANSFYVYQYVDENGKPI-----ENMEFYDRNGN--GTID 874
 QY 859 PADMKVKAANG-----CYLDQFSKPSMR 882
 Db 875 SGDKYIYKRPAGDVLMLGTLKSKQYKNFDFSLR 908

RESULT 9
 Q46034
 ID Q46034

PRELIMINARY; PRT; 2367 AA.

AC Q46034;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Toxin B.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium
 OC Clostridium
 OX NCBI_TaxID=1496;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=Isolate 1470;
 RC MEDLINE=96079281; PubMed=7494480;
 RA von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,
 RA Sattigen S.;
 RT *Closing in on the toxic domain through analysis of a variant
 RT Clostridium difficile cytotoxin B*;
 RL Mol. Microbiol. 17:313-321(1995).
 DR EMBL; 223277; CAAB0815.1;
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR001950; TIF_SUII.
 DR Pfam; PF01473; CW_binding_1; 18.
 DR PROSITE; PS01118; SUII_1; 1.
 SQ SEQUENCE 2367 AA; 269186 MW; EF9823DAE70427F3 CRC64;
 Query Match 3.4%; Score 159.5; DB 2; Length 2367;
 Best Local Similarity 18.4%; Pred. No. 0.45;
 Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;
 QY 27 NSAIYFNTSOPINDQGLAAEVFAQSILPAHPKEDGSOPLHLSLRKSLILLVRPVKAD 86
 Db 1454 NSELQKNIPYSFVDSGK-----ENGFSNGSTKEG-----LFVSELPDVLISKVYMD 1501
 QY 87 DKTP-----VOVEARDNNKILG-----TLTLYPPSSLPDIYHLDGV 124
 Db 1502 DSKPSFGYNNLDAVITKDNVILTYLKYDDIKISLSLTQDEKTIKLSNVHLD-- 1559
 QY 125 PEGGIDFTPHNGTKKIIINTVAAEVNKLSDAGSSSHLTHNALVEIHTANGRWTRDYLP 184
 Db 1560 -ESG-----VAETLKFNKRGS---TNTSDLSMFLSMN---IKSIFV- 1596
 QY 185 QGPDEKGVRFVSSAGY---SSIVFYG-----DRKVTLSVNGT 220
 Db 1597 ---NFGASNKILFDANFIISGTSIGQFEFCDENNNIQPYFIKNTLETNTLYVGNR 1653
 QY 221 L-LFKYNGQWFRSGELENNRITYAQ-HWSAELPAHWIYPGLNLVTKGNLSG----- 272
 Db 1654 QNMIVEPNYDLDDSGDSSIVNFPSQKLYGIDSCYNKVVISPNYITDEINITYETNN 1713
 QY 273 -----RLNDIKI-----GAPGELLHTIDIGMLTTPRDRF-DFAK 306
 Db 1714 TYPEVIVLDANYINEKINVNINDLSIRYVWSNDGNDFILMSTSEENKVSQVIRFVNVFK 1773
 QY 307 DKE-AHREYF---QIPVSRMIVNNYAPLHKEVNL----- 338
 Db 1774 DKTLANKLSFNFSQDQVPVSEIIL-SFTPSYEDGLIGYDGLGLVSLYNEKFIYNNFGMM 1832
 QY 339 -----PTGELLTD-----MDPFGNGHSGTMRQIRIGELVSHGI 372
 Db 1833 VSGLIYINDSLIYFKPPVNNLITGFTVGGDKYFNPINGGAAS-----IGETI---I 1882
 QY 373 DNANYGLNSTAGLNSHPYVVAOLAAHNSRGNY-----ANGIQVHGGSGGGIVTLSDT 427
 Db 1883 DDKNYTFNQS-----VLQTVGFSTEDGKVFAPAN-----TLDEN 1918
 QY 428 LGNE---FSHEVGHNYGLGHYVDGFKGSVHRSANNSTWGW---DQDKRFPINFPSPQ 481
 Db 1919 LEGAIDFTGKLIIDENIYFYEDNYRGAVE-----WKELDEMHIYFSP----- 1962
 QY 482 TNEKSCNLNQCQEPDGKHGFGDAMAGGSPFSAANRF--TMYTPNSSAIIQREFE---EN 535
 Db 1963 -----TGKAFGLNLQIGDDKYFNSDGVNKGQFVSINDN 1996

Qy	189	----	LGKMWRFVSSGAGSYSTVFYGDRTVLVSQNTLLFYVNGQWFRSGELENNRITYAQ	245			
		: : :	: : :	: : :			
Db	504	DKIKSQKISFIKTEEEYINMLKNDKDNVPI-NTLNKSNNNNNNNNNKICSSNNVSPAD	562				
		: : :	: : :	: : :			
Qy	246	HWSAELPAHWIVPCLNLVIRQGNLSRLANDIKIGAPGELLHTIDIGMLTPDRDFDA	305				
		: : :	: : :	: : :			
Db	563	NIFS-----NKSYYDNTKI-----YSEDYINIEKQKEI--FL	594				
		: : :	: : :	: : :			
Qy	306	KDKEA-----HREYFQTI-----VSRMIYNNY-----APLHLKEVMLPTGELLT	345				
		: : :	: : :	: : :			
Db	595	QDOEKENTMNRVRRKESIPNKINDDIVKSDQYWNISNNYEHLISSISKKENENQKRYMD	654				
		: : :	: : :	: : :			
Qy	346	DMDPGNGHWSCTHRO-RIGKELYSHG-----IDNANYGLN--STAGL	385				
		: : :	: : :	: : :			
Db	655	NOMGNNTLHSETFTDIRKDKHSYHNNTCDNIVTEKKNDILHNHNRVINEENIPQI	714				
		: : :	: : :	: : :			
Qy	386	GENSHIPY-----VAQLAAHNSRGNYANGIO-VHGGSGG	418				
		: : :	: : :	: : :			
Db	715	YFNHPPFINVMKRSMSLCNLKDTQVDIDIKVDDLIQVNVHNIYDMYINOKKNIKNSD-	773				
		: : :	: : :	: : :			
Qy	419	GGIVILDSTLGNFSEHVGHYGLHYDGFKESVHSAENNNSTWGDGDKRIPNFI	478				
		: : :	: : :	: : :			
Db	774	---ILFDHTKGLFEN-IKEN---GFYIGPTTTCOIHNPKNND-----INNII	815				
		: : :	: : :	: : :			
Qy	479	PSOTNEKSLCNNOCEPF-----DGHKGFCDAMAGGSPFSAANRFTMYTPNSSAILOR	531				
		: : :	: : :	: : :			
Db	816	---SNNNICEYRKCTSNQNSICYNDQHKLNYYDMGQ-----LNYNPTSKNNMK-	861				
		: : :	: : :	: : :			
Qy	532	FFENKAVTDSRSSTGFSKWNADTOHEPEYHTIDRAEQIT-----ASVNELSEKMAE	584				
		: : :	: : :	: : :			
Db	862	-FTNK-----ENDTNKINIDINNVDNKKSCDDIYYVNMKKNILDNQNNNSOIMLD	913				
		: : :	: : :	: : :			
Qy	585	LMAEYAVVVKHMMNGN-----WTRNIY-----IPTASADNRGSILTIINH-	623				
		: : :	: : :	: : :			
Db	914	IPS-----CHINVKNNNIYDNNLYDNNIYDNNIYDNNKLYDNNSSGVLISHVLSKNNSN	969				
		: : :	: : :	: : :			
Qy	624	-EAGYSNYLPIFNGDEK---VVSQGYKKSFTSDGQFWKERDVT-----REARPEQFG	673				
		: : :	: : :	: : :			
Db	970	NOENYENYKMKNEKKNIYIDMYDQENLMEGKVAEKLILSTLAKNINHKLSETADFG	1026				
		: : :	: : :	: : :			
Qy	674	VPVTVIGVYDPEGLTSSVIYPAMYGAVGYTSDSQNSLSDNDCOL--OVDTKEGQLRFR	731				
		: : :	: : :	: : :			
Db	1030	YKENDLMESKTSNDNTIKIVNSLY-----NDNDCSLINDLNLKQYWDYI	1079				
		: : :	: : :	: : :			
Qy	732	LANHRANNTVMKFNHIN---VPTESQPTQATLVCN-NKILDTKSUTPAPEGLTYTVNGQ	786				
		: : :	: : :	: : :			
Db	1076	MNN---KNNVNNKNNVNNMNSVNNNYYNNNNNNNNNNVNTYSLN-----NTK	1122				
		: : :	: : :	: : :			
Qy	787	ALPAKENEGCIVSVNSGRKRYCLPVQORSYS--LPDWI-----VQEVYVDSG---AKA	835				
		: : :	: : :	: : :			
Db	1123	NIDNTQDSKLSVMSFNTFCOMKDEFFNOFLEHFKINSNNVVEEILNDQKNEILKN	1188				
		: : :	: : :	: : :			
Qy	836	KVLLSOWDNLN--YNRIGEFVGNVPADMK	863				
		: : :	: : :	: : :			
Db	1183	KIIISDINNKEVYNKONE--NNNNNLSVK	1209				
		: : :	: : :	: : :			
RESULT 11							
Q37F62							
ID	Q37F62	PRELIMINARY;	PRT;	1157 AA.			
AC	Q37F62;						
DT	01-OCT-2001 (TEMBLrel. 18, Created)						
DT	01-OCT-2001 (TEMBLrel. 18, Last sequence update)						
DT	01-MAR-2003 (TEMBLrel. 23, Last annotation update)						
DE	Fusion of alpha-glucosidase (Family 31 glycosyl hydrolase) and						
DE	glycosidase (TreA/Mais family).						
GN	CAC2891.						
OS	Clostridium acetobutylicum.						
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;						
OC	Clostridium.						
OX	NCBI_Taxid=1488;						
RP	SEQUENCE FROM N.A.						
	[1]						

RC STRAIN-ATCC 824 / DSM 792 / VEM B-1787;
RX MEDLINE-21359325; PubMed-11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007786; AAK80833.1;
DR InterPro: IPR005085; CBM_25.
DR InterPro: IPR000322; Glyco_hydro_31.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF03423; CBM_25; 1.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
DR Hydrolase; Complete proteome.
SQ SEQUENCE 1157 AA; 129411 MW; 1DB1728D833F7578 CRC64;

Query Match 3.3%; Score 156.5; DB 16; Length 1157;
Best Local Similarity 19.3%; Pred. No. 0.23;
Matches 198; Conservative 120; Mismatches 335; Indels 371; Gaps 56;

QY 83 VKADRTKPYOVVEARDNNK-----ILGTLTLPSSLPDIYHLCQVPEGIDFTPHNCTK 138
DB 35 VKADTKVNVKNKSSQSKFKAKLNGTLKIKGKDETLIRICEPQVFKVDYKP-NGKS 93
QY 139 KIINTVAEYNKLSDSGSIHSLTNALVEIHTANGWRDI-----YLPQGPDL 189
DB 94 SKDTLVVDPNK-KWSTGNIVSDIISDPWV---ITTKMKVLKINKEDLSILVY-----DL 144
QY 190 EGMWRFVSAGYSSTVF---YCDKRVTLVSGNTLLFKYVNGQWFR-----SGELEN-NRI 241
DB 145 OGRLLKQDSTAKTSASFTHNSGDR-----FYGINGYFQEDSSKGLMRNGTES 193
QY 242 TYAQHTWSAELPAHVIVPGLNVIKGNLSGRLND-----IKIGAPGEL 285
DB 194 VTAGYGHGCGSPVWNSDGLLVSDSGSFTIGTSLKYDGIKTDYDYYVMGNPKREI 253
QY 286 LLHTIDIG-----MLTTPDRF-----DFADKEAHREY-FQTIPVSRMI 324
DB 254 LSESDVSGKAPMPKAWNGFTNTQWGNLSLGGTNDKALKLSINTYRSKQLPIDNFC 313
QY 325 V-----NNYAPLHLKEVMP--TCELLTMDPGNGHSGTMRQRI-----GKE 366
DB 314 LDFDKWKWGDNYGEFKWNTDNPDSQGLKAYMD-SKGLKMTGIMKPRILADSKQGRY 372
QY 367 LVSHG-----IDNANYGLNSTAGLGENSEHPYVVAQLAAHNSRGNANG 409
DB 373 VTSKGNWLPDSEASDYCSCKMKNVNFAL-----PQV-----RKWNNN 412
QY 410 IQVHGSGGGGIVTLDSTLGNESHEVGHNYGLGHY-----VDGFKGSHRSANN 460
DB 413 IQ---GAFDKGIVGF-----WNDCDENVNFNGNMMERAIYDG-----ORRHKN 456
QY 461 NSTWG-----WDGDKKRTIPNFPSTNEK-----SCLNQCQEPFDGKH 500
DB 457 QRWVSLNRNYAQAQRYSGVMWSD-----ISTGFSMAQRERMLSAVN-----LGEAK 506
QY 501 FGFD--AMAGSP-----FSAANRFTWYTPNSSAIIQRF--ENKA-----VFSRSSTGFS 548
DB 507 WGMTDGGFGDPTPENYARWMEFSAFTP-----IPRVHGQDNKYRYPWAFGSTAEATAK 561
QY 549 KWNADTOEMEPYEHITDRAQITA-----SVNELSEKMAELMAYAVYKVKM 596
DB 562 KAMQLRTPLIPIYSYDRSAQSGLGLVRLSLAMXPNDNAANDKAWMFGDMLVSPVY 621
QY 597 WNGWTRNIYIP-----TASADNRGSLITINHEAGNSY---LFING----- 635
DB 622 QEGQTSKSIITLPGNWDIYTGREYTG--QTINAVDSTNWDSDIPIFKSGAIITQDFE 680
QY 636 ---DEKVVSGYKKSF-----VSDGQFKWERDVIDTREARKPEQFG 673
DB 636 ---DEKVVSGYKKSF-----VSDGQFKWERDVIDTREARKPEQFG 673

DB 681 NYVGEKKITDYVDAPPNGNEASSFDYDDGTSYNYENGSYFDQKMTLERADKLKSVQFN 740
QY 674 VPVTTLVGYDPEGTLSSYIP-----AMYGAYGFTYSDQSQNLSDNDC 717
DB 741 ISPKT--GY--KSDLKNYIVKMHVKSSGDVTVGGRITRYASY-----DELKNAQGEY 791
QY 718 QLOVDTKEGQLAPRL-ANHRANNTVMKFNHINVPTESQPTQATLVCNNKIL-----DTKSL 772
DB 792 VGTDTYGSVYIKVSAGHDKN-----INP-----CNQVLTAYADVK-- 830
QY 773 TPAPEGLTYTVAQGALPAKEN-----EGCIVSVNSGKRYCLPVQSRGYSLPDWI 822
DB 831 -----GGTYT-SPQKVSLSKASDPNAAIYVTLDTAPTAVNSTK-----YTGP--- 870
QY 823 VGOEYTVDSGAKAKVLLSDWNLNSNRIGEFVGNVNPADKKVKAANGQYLDFSKPRSMR 882
DB 871 -----ITIDSKTLKFIVRD-----ANGNESDV-----FTEQYTTY-----IK 903
QY 883 VVYK 886
DB 904 VHYK 907

RESULT 12
QYF931
ID QYF931 PRELIMINARY; PRT; 2367 AA.
AC QYF931;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DB Cytotoxin B.
CN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID:1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-5340;
RX MEDLINE-20448897; PubMed-10992443;
RA Sambol S.P., Merrigan M.M., Lierly D., Gerding D.N., Johnson S.;
RT "Toxin gene analysis of a variant strain of clostridium difficile that
RT causes human clinical disease";
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL: AF217292; AAG18011.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR001950; TIF_SUII.
DR Pfam: PF01473; CW_binding.1; 18.
DR PROSITE: PS01118; SUII.1; 1.
SQ SEQUENCE 2367 AA; 269337 MW; D5EE715E5BD41E2F CRC64;

Query Match 3.3%; Score 154.5; DB 2; Length 2367;
Best Local Similarity 18.4%; Pred. No. 0.95;
Matches 185; Conservative 148; Mismatches 322; Indels 353; Gaps 49;

QY 27 NSAIYNTSQPINDLQGLAAEYKFAQSQILPAHPREGDSQPHLTSLRSLSLVPRVKAD 86
DB 1454 NSELQRNIPSYFVDESK-----ENFGSTKEG-----LFSVSELPDVVLISKVYND 1501
QY 87 DKTP-----VQVEARDNNKILG-----TLTLPSSLPDIYHLDGV 124
DB 1502 DSKPSFCYYSNNLKDVKITKQNVNLTGYYLKDDIKISLTLQDEKTIKLSVHLD-- 1559
QY 125 PEGGIDTPPINGTKKINTVAEYVKLSADSGSIHSLTNALVEIHTANGWRDIYLP 184
DB 1560 -ESC-----VAETLKFPNRRKGS---TNTSDLSMFSLESNM---IKSIFV- 1596
QY 185 QGPDLECKMVRFYSSAGY---SSTVFYG-----DRKVTLSVGNT 220
DB 1597 ---NFLQSNKIFLDANFIISGTISGQFEICDENNNIOPIFKNTLETNTLYVGNR 1653
QY 221 L-LFKYVNGOWFRSGELENRRITYAQ-HWMSAELPAHVIVPGLNVLKQCNLSG----- 272

Db 1654 QNIVPEPNYDLDDSGDISSTVINPQKYLIGIDSCVNKVVISPNIYTDIENITPVYETNN 1713

Qy 273 : : : : : RNDIKI- : : : : : GAPCELLHTIDIGMLTTTPDRF-DFAK 306

Db 1714 : : : : : INEINVDLSIRYVNSDGNDFILMSTSEENKVSQKIRVNVFK 1773

Qy 307 : : : : : QTPVSRMIVNNYAPLHLKEVNL- : : : : : 338

Db 1774 : : : : : DKTLANKLSFNSDKQDVPVSEIL-SETPSYEDGLIGYDGLVSLVNEKFIYNNFOMM 1832

Qy 339 : : : : : PTELTD- : : : : : MDPGNGHSGTMRQRIKELVSHGI 372

Db 1833 : : : : : VSLIIVINDSLYFKPPVNNLITGFVVGDDKYFENPINGGAAS- : : : : : IGETI- : : : : : I 1882

Qy 373 : : : : : DNANYGLNSTAGLGENSHPYVVAQLAAHNSRNY- : : : : : ANGIQVHGSGGGGIVTLDTST 427

Db 1883 : : : : : DKNYFVNSG- : : : : : VLOTGVFSTEDGKYFAPAN- : : : : : TLDEN 1918

Qy 428 : : : : : LGNE- : : : : : FSEVGHNYGLGHYDGFKGSVHRSANNNSTGW- : : : : : DGDKKRFPNFPYFSQ 481

Db 1919 : : : : : LEGEADFTGKLIIDENIYFEDNYGAVE- : : : : : WKLDGEMHYFSP- : : : : : 1962

Qy 482 : : : : : TNEKSLNNQCQPPDGHKFGFDAMAGSPFSAANRF- : : : : : TMYTPNSSAIQRF- : : : : : EN 535

Db 1963 : : : : : : : : : : : TGAFKGLNIGDDKYYFNSDGMQKGFVSINDN 1996

Qy 536 : : : : : KAVFDSRS- : : : : : STGFSK- : : : : : WADTQEMEPYEHTIDRAOITASVNELSRAELMA 587

Db 1997 : : : : : KHYFDDSGVMKVGYTEIDGKHFFAFENGEMOIGVFNTDGFYFAHNNEDLNGEGERIS 2056

Qy 588 : : : : : EYAVVYVHMNGWNTNIYIPTASADNRGSIITINHEAGNSYLFINGDEKVVSGYKKS 647

Db 2057 : : : : : YSGIL- : : : : : NFNKIY- : : : : : FDSFTAVVGVCKDLEDGSKYF- : : : : : DEDTAEAYIGLS 2103

Qy 648 : : : : : FVSDGQFWKROVVDREARKPQFG-VPVTLVGYDPEGTLSSYIYPAMYGATFYYS 706

Db 2104 : : : : : LINDGQYFNDG- : : : : : MQGVFTINDKVFYFSDSGIIES- : : : : : GVONI 2145

Qy 707 : : : : : DQONLSDNDCOLQV- : : : : : DTKEGOLAFRLANRANNTVANKFHNVPYTESQPTQATLVGN 763

Db 2146 : : : : : DDNYFIDDDNGIYQIGVFTSDGKYFAPA- : : : : : NTVNDNIY- : : : : : 2183

Qy 764 : : : : : NKILDTKSLTPAPE- : : : : : GLTYVNGOALPAKENECCIVSNVSGKRYCLPVGQRS- : : : : : GY 816

Db 2184 : : : : : QAVEXSGLVRVEDVYFGETYTIETGWIDMENE- : : : : : SDKYYFDPETKACKGI 2236

Qy 817 : : : : : SLPDWIVGOEYVYDSCAKAKVLLSDMDLSY- : : : : : NRIGEF-VGVNVPAD 861

Db 2237 : : : : : NLIDDI- : : : : : KYFDEKGMRTGLISFENNYFNENGEMQFGYINIED 2281

RESULT 13

Q8KRR3 PRELIMINARY; PRT: 3692 AA.

AC Q8KRR3, 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Outer membrane protein.

OS Fusobacterium nucleatum.

OC Bacteria; Fusobacteria, Fusobacteriales; Fusobacteriaceae;

OC Fusobacterium.

OX NCBI_TaxID=851;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PK1594;

RA Ma X., Kempf M.J., Jewett A., Park H.-H., Shi W.;

RT "Cloning and analysis of Fusobacterium nucleatum apoptosis-inducing genes";

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF525505; AAM09095.1; -

DR InterPro; IPR005546; Autoporter.

DR Pfam; PF03797; Autoporter; 1.

SO SEQUENCE 3692 AA; 389826 MW; DEF61C008A0ED180 CRC64;

Query Match 3.2%; Score 150; DB 2; Length 3692;

Best local similarity 18.9%; Pred. No. 3.8;

Matches 188; Conservative 113; Mismatches 358; Indels 338; Gaps 42;

Qy 10 ILAPLAIGVFSATAADNNSAIVENTSQPINDDQSLAAEVKFAQSOI- : : : : : LPAHKEGDS 66

Db 92 IETPTVNVAPPVAAIN-PVAFSVAPTIDSKQ- : : : : : YKFGGANINGLAQLSTIDV 144

Qy 67 QPHLTSLRKSLLL- : : : : : VREPKADDKTPVQVEARDNNKIL- : : : : : GT 105

Db 145 QSNSTNRNTVTTINASSGVSTSVLLPDKKTINVS- : : : : : NNNRALVYDEARENFDFQMGKT 202

Qy 106 LTLYPSSSLPDTIY- : : : : : HLDGVPEGGID- : : : : : FTHNGTKKLIINTVAEVNKL- 150

Db 203 INLYGNQNMGIDLOGTHIGCSAKGSATNPAISTIRNEGVTGH-ATPNKYDNKATKEOVA 261

Qy 151 : : : : : SDAGSSITSHLTHLNNALVEIHTANGRWYRDYLPQGPDLCKGMVFVSSAGYST 205

Db 262 FGSNADASSNATWTHNNNEITLAPSSAGIQ- : : : : : LKPEDP- : : : : : 301

Qy 206 VFYGDREK-VTLSVGNLTLLFKYVNGOMFRSGELENNRITYAQHITWSAELPAHMIYVPGNLIV 264

Db 302 -FYMDPNWGTLSNDKMIINGLSG- : : : : : SRNANNEGRVLMKADNR- : : : : : KDINL- 346

Qy 265 IKQNLISRLNDIKIGAPGELLHTIDIGMLTTPROFDFAKDEAHREYFQTIPVSRMI 324

Db 347 : : : : : NGSGFGMITVFPVPGVIELTISVNRITPTGSI- : : : : : 380

Qy 325 VNNYAPLHLKEVMLPTGEL- : : : : : LTDMDPGNGHSGTMRQRIKELVSHGI- : : : : : DNANYG 378

Db 381 -NLRAQRNIGTKVLPGEIGRSALSDSKTSYGVNSGNININ-GDESVGVLHLEIOEVK 438

Qy 379 LNSTAGLGENSHPYVVAQLAAHNSRNGYANGIOVHGSGGGGIVTL- : : : : : STLCNE- 431

Db 439 IGGTINIGSVSQTGISTKSTGNDQDTLLVNNAVGVFSGYPTLFPKNEYDTHGNKN 493

Qy 432 : : : : : FSEVGHNYGLG-HYVDGFKGSVHRSAN- : : : : : NNSTGWGDDGKKRFPN- : : : : : F 477

Db 499 TTGNLIGTSEVDTGNTIGRHSKESICLLVDSGETLNGTNGTNPQARILKRGSSIT 558

Qy 478 YPSQTNKESCLNNQCQPPF- : : : : : DGHK- : : : : : 500

Db 559 YKSSANNKLNINGNANYGFVYKSDSKSVFCSALDDLQOSVDKNTYNGIGINKNIDIIGI 618

Qy 501 : : : : : PGFDAMAGSPFSAAN- : : : : : RFT- : : : : : 519

Db 619 NSIGFVLLKGNKNTGNTIVNESEDFWYPIVPTPNYDYRQNSIGFYGQDNFTNEGTI 678

Qy 520 -MYTPNSSAIQRFENKAVFDSRSSTGFSKWNADTQEME- : : : : : PYEHTID 565

Db 679 SVNTPNRSG- : : : : : NKAVLLKGNNGITFNNTGDSVKGRNNIAIYABGKYTFNH- : : : : : 728

Qy 566 RAEQITASVNELSESKMAELMAEYAVVYVHMNGWNTNIYIPTASADN- : : : : : 614

Db 729 -EKNAAGTNKISVGSNS- : : : : : IGIVYKONTGTVINIKAPIELADSSNGTIGVYSD 776

Qy 615 : : : : : RGSILTINHEA- : : : : : GYNSYLFIN- : : : : : GDEKVVSGQYKKSFGVSDGF 654

Db 779 GNAHINFGSKLTIGKKAIGLFSQSTTNFKNTFKFNTAGNELNVLSDNSAFA- : : : : : F 834

Qy 655 WKERDVVDTREA- : : : : : RKPEQF- : : : : : GVPVTLVGYDPEGTLSYIYPAMYGA 700

Db 835 FNGSGSTDIAEVLNKNIKFTAMKKGATFAYVKGGSIVTLSDHEDFTTNTAKVYVAPES-GT 895

Qy 701 YGFYISDSDQLSDNDCOLQVDTKEGQLR- : : : : : 735

Db 894 SVLVASDGSITVOVDANKKLTNTNVGLVATNGSSSKSKAINKGTIISKVNGVGGLYANN 953

Qy 734 NHRANNTVNNKFHNVPYTESQPTQATLVCNKILDTK 770

Db 954 NSEATNDIAGK- : : : : : ITHENKGSAILGENNSILTNK 986

Db 532 MDTGFGNQQDGTNNPNDLYARMQFSALPVPVFRVGNHNNHQRPWFSGTAEASKEA 591

QY 568 EQ1-----TASYNELSEKMAELMAEY---AVVK--VHMWN-GNW-----601

Db 592 IQLRSLIPYMYAYERSAYENGGLVRLMQAYPTDAVKNYTDAWFGDWLLAAPVVK 651

QY 602 ---TRNIYIPTASDN--RGSILT-----INHEAGYNSYLFINGDEKVKVSGYKKS 648

Db 652 QOTSNDIYLPSCSWIDYARGNAITGGOTIRYSVNPDTLTOMPLFI-----KKGA 700

QY 649 VSDGQFWKERVVDTRARKPEQFGVPTTLVG--YYDEPTLSSYIIPAMYGAIGFTYS 706

Db 701 IIPQ--KVQDYVGOASVTSVDVDVFPDTTQSSFTYYDDG-----ASYNVESCTYF 750

QY 707 DDSQNLSDNDQOLQVDTREGQLRFLANHRANNTVMKFEHINVTESQPTQATLVCNKI 766

Db 751 KQMTAQDNG-----SGLSFTLGAKSGSYTPALQSYI---VKLHSGAGTSVTNNS- 798

QY 767 LDTKSLTTPAPEGLTYTVNGQALPAKNEGCVIVSVNKGKRYCLPVGORSYSLPDWIVQGE 826

Db 799 -----AASTSYASLEALKAAAGES-----WATGKD 823

QY 827 VTVD 830

Db 824 IYGD 827

RESULT 15

O912M3 PRELIMINARY; PRT; 2468 AA.

AC Q912M3; PRELIMINARY; PRT; 2468 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Hypothetical protein PA1874.

GN PA1874.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxId=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.I., Erwin A.L., Miziochichi S.D., Warrenner P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V., "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";

RT Nature 406:959-964 (2000).

RL EMBL; AB004613; AAG05263.1; -.

DR InterPro; IPR001343; Hemlysn_Ca_bind.

DR InterPro; IPR006162; Ppantne_attach.

DR PRINTS; PR003113; CABNDMGRT.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 2468 AA; 238414 MW; 13596AFAB2C4B899 CRC64;

Query Match 3.2%; Score 148.5; DB 16; Length 2468;

Best Local Similarity 19.5%; Pred. No. 2.5;

Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 33 NTSQINDIQLGSLAAEFKFAQSQILP-----AHP-----KEGDSQPH-----69

Db 324 NNSTPVTVEAPOTTAPATDQVAPDGSSVTGNAEPGATGVDTDGQPDVTWVGP 383

QY 70 -----LTSLRKSLILVPRVADDTKPTQVVEARD-----DNNKIL-----103

Db 384 GSPEVPLNPLNGEFTVITVDPAGNCSSTPVTAPDFPDPAPQVNASNGSVLSTGAEAG 443

QY 104 -----CTHLYPPSSLPD-----TIYHLDGV 124
Db : : : : :
444 VTIVIDGNGNPIGOTSADANGNSFTQSQPLDGTVVNVYVARDAGNSSPATSTIVDGV 503
QY 125 PEGGIDFTPHNGTKKIINTVAEVNK---LSDASGSIHSLTNNALVIEIHTANGRWVRDI 181
Db : : : : :
504 APNAPVVEPSNGSE--LSGTAEPGSSVTLTDGNGNPIGOTTAD-----ANGNW---S 550
QY 182 YLPQGPDLGSKKVRVFS--SAGYSS-----TVFYGDRKVTLSVGNLTLLFKYVNGOMF 231
Db : : : : :
551 FTFSTPLPDGTVVNVYVARDAGNSPPASVTVDAVAPATPTVDPNSGTTL----- 600
QY 232 RSGELENNRITYAQHIWSAELPAHMIIVPGLNLVIRKQNI/SGRLNDIKIKAPCELLLHTID 291
Db : : : : :
601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 292 IGMLTTPRDRDFEAKDEHREYFQTIPIVSRMIVVNYAPLHLKEVMLPTGELLTD----- 346
Db : : : : :
620 IGQVTA-----DGSNGMTFTFTPLPNCVTVNATAD-----PSGNASSPASVTV 664
QY 347 -----MDPGNGWHSGTMRQIRIGKELVSHGIDNANYGLNSTAGLGENSHPYVVAOL 397
Db : : : : :
665 DAVAPATPVVNSNGTTLSGT-----AEPGATVTLTDG-NGNP--ICQV 705
QY 398 AAHNSRGNYA-----NGIQVH-----GGSGGGGIVTLDS-----TL 428
Db : : : : :
706 TADGS-GNMSFTPTPLPNCVTVNATATDASGNTSAGSSVTVDSVAPATPVINPSNGTTL 764
QY 429 GNEFSHEVGHNTGLHYVDGFKGSHVRSFAENNSNTHGWDGDKKREIPNFPYS-OTNEKSC 487
Db : : : : :
765 SG--TAEPGSSVTL---TDGNGNP/IGQVTDGSGNW-----SFTPTPLADGTV 808
QY 488 LNNQCOEPFDGHKFGPDAMAGGSPESAANRFTWYTPNSSAI/IRFFENKAVFDSRSSTGF 547
Db : : : : :
809 VNATATDP-----AGNTSGGSTVDGVAPTPTV-----NLSNGSSLGTA- 850
QY 548 SKWNADTQEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVVYVHMNONGNWTNRNIYI 607
Db : : : : :
851 -----EPGS-----TVILTDGNGNPIAEVTAD-----GSGNWT---YT 880
QY 608 PTASADNRGSILTIHNEAGVNSYLFINGDEKVVYSGYKKSFYSDGQFWKRDVVDTREAR 667
Db : : : : :
881 PSTPIANGIVNVVVAODAAGNS-----SPGASV-----VDSQAPA 916
QY 668 KPEQFGVPVTVTLVGYDPEGTLSSYIYPAMYGAYGFTYSDDSQNLSDNDCLQVDTKEGQ 727
Db : : : : :
917 APVVPNSNGTTLSGTAEPGATV-----TLTDGNGNPIG---QVTAD-GSGN 958
QY 728 LRFLANHRANNTVMKPHINVPTESQPTOATLVCNNKILDTKSLTPAPEGLTYTVNGOA 787
Db : : : : :
959 WSFTPGTPLANGTVWN-----ATASDPTGNTSAPASTTVD--SVAPAAP-VVNPNSNGAE 1009
QY 788 LPAKENEGCIVSVNSGKRYCLPVGQRSYSLPDW 821
Db : : : : :
1010 ISGTAEPGATVTLTDGSGN--PIGQVTDGSGNW 1041

Search completed: October 8, 2003, 16:27:19
Job time : 54.5173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 8, 2003, 16:20:57 ; Search time 21.2134 Seconds
(without alignments)
3912.317 Million cell updates/sec

Title: US-10-002-309b-2_COPY_24_886
Perfect score: 4597
Sequence: 1 ADNSAIYFNTSQPINDLQG.....AWNGYLDKSPRSMRVYK 863

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4597	100.0	898	2 T42131	probable toxR-regu
2	2604.5	56.7	587	2 T00316	toxR-regulated lip
3	689.5	15.0	1002	2 T09438	toxR-activated lip
4	689.5	15.0	1013	2 B82276	ToxR-activated gen
5	671	14.6	1335	2 A82494	TagA-related prote
6	290	6.3	937	2 H82261	hemolysin-related
7	166.5	3.6	1217	2 F97177	alpha-glucosidase
8	161	3.5	1038	2 JC6027	115K outer membran
9	159.5	3.5	2367	2 S70172	toxin B - Clostrid
10	156.5	3.4	1157	2 F97255	fusion of alpha-gl
11	154	3.4	1211	2 S54500	alpha,alpha-trehal
12	153.5	3.3	2366	2 S10317	toxin B - Clostrid
13	148.5	3.2	2488	2 A83412	hypothetical prote
14	148	3.2	1021	2 I40805	collagenase - Clos
15	147	3.2	1386	2 AC1533	surface protein (L
16	146.5	3.2	1377	2 D90538	hypothetical prote
17	139	3.0	1530	2 AH1396	peptidoglycan anch
18	138.5	3.0	1268	2 B99789	hemagglutinin/hemo
19	138.5	3.0	1270	2 E85649	hypothetical prote
20	138	3.0	2894	2 C64474	hypothetical prote
21	136.5	3.0	973	2 B86547	polymorphic outer
22	136.5	3.0	973	2 F72076	polymorphic outer
23	136.5	3.0	995	2 C81593	polymorphic membra
24	136.5	3.0	3705	2 AD0123	probable autotrans
25	136	3.0	2167	2 AF1489	cell wall-associat
26	135.5	2.9	2529	2 B64635	toxin-like outer m
27	135	2.9	1579	2 B91290	probable invasin I
28	135	2.9	1700	2 G86131	probable invasin 2
29	134.5	2.9	1026	1 TLB774	tail fiber protein

30	133.5	2.9	2660	2 E85822	probable invasin 2
31	133	2.9	4199	2 S76412	hypothetical prote
32	132	2.9	815	2 H83345	ferrityoverdine re
33	132	2.9	2273	2 T09083	hemagglutinin/hemo
34	130.5	2.8	773	2 JE0387	exo-alpha-sialidas
35	130	2.8	813	2 A40601	ferrityoverdine re
36	129.5	2.8	2186	2 T13169	tiggrin - fruit fl
37	129	2.8	1250	2 D91018	hypothetical prote
38	128.5	2.8	1004	2 B25039	outer cell wall pr
39	128.5	2.8	1051	2 G84316	ribonucleoside red
40	128.5	2.8	1238	2 A64596	hypothetical prote
41	128.5	2.8	1939	2 D97316	probable S-layer p
42	128	2.8	1250	2 F85862	hypothetical prote
43	128	2.8	1487	2 AG2560	hypothetical prote
44	127.5	2.8	936	2 S57637	hexon protein - hu
45	127.5	2.8	1067	2 D75625	probable extracell

ALIGNMENTS

RESULT 1

T42131
probable toxR-regulated lipoprotein tagA - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C:Accession: T42131; T00210
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Esc
A:Reference number: 222068; MUID:98391744; PMID:9722840
A:Accession: T42131
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-898 <BUR>
A:Cross-references: EMBL:AF074613; PIDN:AAC70099.1
A:Experimental source: strain EDL933; serotype O157:H7
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo
S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemor
A:Reference number: 214127; MUID:98290540; PMID:9628576
A:Accession: T00210
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 569-898 <MAK>
A:Cross-references: EMBL:AB011549; NID:d1204561; PIDN:BAA31757.1; PID:d1032718
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952
C:Genetics:
A:Genome: plasmid p0157
A:Note: L7031

Query Match
Best Local Similarity 100.0%; Score 4597; DB 2; Length 898;
Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ADNSAIYFNTSQPINDLQSLAAEYKFAQSILPAHPKEDSQPHLTSLRKSLLLRVPY 60		
DB	36	ADNSAIYFNTSQPINDLQSLAAEYKFAQSILPAHPKEDSQPHLTSLRKSLLLRVPY 95		
QY	61	KADDKTPQVQEARDDNNKILGTLTLYPPSSLPTDIYHLDGVPGGIDFTPHNGTKKIINT 120		
DB	96	KADDKTPQVQEARDDNNKILGTLTLYPPSSLPTDIYHLDGVPGGIDFTPHNGTKKIINT 155		
QY	121	VAEYKLSDSAGSGSIHSHLTNNALVEIHTANGRWVRDIYLPQGDPLEGKGRVFSAGYS 180		
DB	156	VAEYKLSDSAGSGSIHSHLTNNALVEIHTANGRWVRDIYLPQGDPLEGKGRVFSAGYS 215		
QY	181	STVYIGDKVTLSVGNLTLLFKYNGQWFRSGELNNRITYAHIWISAEPLPAHWIVPGLNL 240		
DB	216	STVYIGDKVTLSVGNLTLLFKYNGQWFRSGELNNRITYAHIWISAEPLPAHWIVPGLNL 275		
QY	241	VIKQNLGSLRNDIKIGAPGELLHTIDIGMLTTPDRDFDFAKDKAHEHYFQTIIPVSRM 300		

|||||
276 VTRQNTLSGRINDIKIGAPGELLHTIDIGMLTPDRDFDAKDKAHEHYFOTIPVSRM 335
QY
301 IVNNYAPLHLKEVMLPTGELLTDMPCGNGHSGTMRQIRIGKELVSHGIDNANYGLNSTA 360
Db
336 IVNNYAPLHLKEVMLPTGELLTDMPCGNGHSGTMRQIRIGKELVSHGIDNANYGLNSTA 395
QY
361 GLGNSHPYVVAQAAHNSRGNTANGIQVHGSGGGGIVTLDSTLGNFSHEVGHYGLG 420
Db
396 GLGNSHPYVVAQAAHNSRGNTANGIQVHGSGGGGIVTLDSTLGNFSHEVGHYGLG 455
QY
421 HYVDFKGSVHRSRNNKNTWGDGKKRIPNFPYQINTEKSCNNOCQEPDGHKFGF 480
Db
456 HYVDFKGSVHRSRNNKNTWGDGKKRIPNFPYQINTEKSCNNOCQEPDGHKFGF 515
QY
481 DAMAGGSPFSAANRFTMYTPNSSAIIQRTFENKAVDSRSTGFSKWNADTQEMPEYHT 540
Db
516 DAMAGGSPFSAANRFTMYTPNSSAIIQRTFENKAVDSRSTGFSKWNADTQEMPEYHT 575
QY
541 IDRAEQITASVNELSEKMAELMAEYAVVVMWNGNWTNIIYIPTASADNRGSILTIH 600
Db
576 IDRAEQITASVNELSEKMAELMAEYAVVVMWNGNWTNIIYIPTASADNRGSILTIH 635
QY
601 EAGNSYLFINGDEKVSQGYKSFVSDGOFWKEKROVVDTRARKPQFGVPVTVLVGY 660
Db
636 EAGNSYLFINGDEKVSQGYKSFVSDGOFWKEKROVVDTRARKPQFGVPVTVLVGY 695
QY
661 DPEGLTSSYIYPAMYGAYGFTYSDSONLSDNDQLOVDTKEGOLRPLANHRANNTVMN 720
Db
696 DPEGLTSSYIYPAMYGAYGFTYSDSONLSDNDQLOVDTKEGOLRPLANHRANNTVMN 755
QY
721 KFHINVTESQTOATLVNCKNIIKDTSLTPAPEGLTYTVNGQALPAKENECCIVSVNSG 780
Db
756 KFHINVTESQTOATLVNCKNIIKDTSLTPAPEGLTYTVNGQALPAKENECCIVSVNSG 815
QY
781 KRYCLPVCORSGYSLPDWIVGOEYVDSGAKAKVLLSDWNLNSYNRIGFVGVNVPADMK 840
Db
816 KRYCLPVCORSGYSLPDWIVGOEYVDSGAKAKVLLSDWNLNSYNRIGFVGVNVPADMK 875
QY
841 KYKAWNGOYLDKSPRSRVRVYK 863
Db
876 KYKAWNGOYLDKSPRSRVRVYK 898
RESULT 2
T00316
toxR-regulated lipoprotein tagA - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 11-Jan-2002
C:Accession: T00316
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota, S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli O157:H7 strain
A:Reference number: Z14127; MUID:98290540; PMID:9628576
A:Accession: T00316
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-587 <MAK>
A:Cross-references: EMBL:AB011549; NID:d1204561; PIDN:BA31835.1
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: tagA
A:Genome: plasmid p0157
Query Match 56.7%; Score 2604.5; DB 2; Length 587;
Best Local Similarity: 96.8%; Pred. No. 3.7e-162;
Matches 491; Conservative 2; Mismatches 7; Indels 7; Gaps 2;
QY 29 AQSQILPAHPKEDGSDPHLTSLRKSLLLVPRVKADDKTPVQVEARDNNKILGTLTLYPP 88
Db 63 AASQILPAHPKEDGSDPHLTSLRKSLLLVPRVKADDKTPVQVEARDNNKILGTLTLYPP 122

89 SSLPDTIYHLDGVPPEGIDFTPHNGTKIINTVAEYVKNKLSDSAGSSSHSLTNNALVEIH 148
Db 123 SSLPDTIYHLDGVPPEGIDFTPHNGTKIINTVAEYVKNKLSDSAGSSSHSLTNNALVEIH 182
QY 149 TANGRWVRDIYLPQPDLEGKMWRFVSSAGYSSTVFYGDGRKVTLSVGNLTLLPKYVANGOWF 208
Db 183 TANGRWVRDIYLPQPDLEGKMWRFVSSAGYSSTVFYGDGRKVTLSVGNLTLLPKYVANGOWF 242
QY 209 RSGELNNRIYTAQHINSAELPAHVIIVPGLNLVTKQNLSGRLNDIKIGAPGELLHTID 268
Db 243 RSGELNNRIYTAQHINSAELPAHVIIVPGLNLVTKQNLSGRLNDIKIGAPGELLHTID 302
QY 269 IGMLTTPDRDFDAKDKAHEHYFOTIPVSRMIVNNTAPLHLKEVMLPTGELLTDMDPGN 328
Db 303 IGMLTTPDRDFDAKDKAHEHYFOTIPVSRMIVNNTAPLHLKEVMLPTGELLTDMDPGN 362
QY 329 GWHSGTMRQIRIGKELVSHGIDNANYGLNSTAGLGENSHBPYVVAQAAHNSRGNYANGIQ 388
Db 363 GWHSGTMRQIRIGKELVSHGIDNANYGLNSTAGLGENSHBPYVVAQAAHNSRGNYANGIQ 422
QY 389 VHGGSGGGIVTLDSTLGNFSHEVGHYGLVVDGFKGSVHRSRNNKNTWGDGDKK 448
Db 423 VHGGSGGGIVTLDSTLGNFSHEVGHYGLVVDGFKGSVHRSRNNKNTWGDGDKK 482
QY 449 RFIPNFPYQINTEKSCNNOCQEPDGHKFGFADAMAGSPFSAANRFTMYTPNSSAIIQ 508
Db 483 RFIPNFPYQINTEKSCNNOCQEPDGHKFGFADAMAGSPFSAANRFTMYTPNSSAIIQ 542
QY 509 FFENKAVF----DSRSSTGFS---KWN 528
Db 543 FFENKAVSIAPPPASAGMQIRKWN 569
RESULT 3
T09438
toxR-activated lipoprotein - Vibrio cholerae
N:Alternate names: toxR-activated gene A protein
C:Species: Vibrio cholerae
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T09438; A39108
R:Karolis, D.K.R.; Johnson, J.A.; Bailey, C.C.; Boedeker, E.C.; Kapur, J.B.; Reeves, Proc. Natl. Acad. Sci. U.S.A. 95, 3134-3139, 1998
A:Title: A Vibrio cholerae pathogenicity island associated with epidemic and pandemic A:Reference number: Z16672; MUID:98169509; PMID:9501228
A:Accession: T09438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1002 <KAR>
A:Cross-references: EMBL:AF034434; NID:g3004923; PIDN:AAC12274.1; PID:g3004926
A:Experimental source: strain N16961
R:Parsot, C.; Taxman, E.; Mekalanos, J.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 1641-1645, 1991
A:Title: ToxR regulates the production of lipoproteins and the expression of serum re A:Reference number: A39108; MUID:91156664; PMID:2000374
A:Accession: A39108
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-25 <PAR>
A:Cross-references: GB:M60658
C:Genetics:
A:Gene: tagA
A:Note: part of the pathogenicity island (VPI); associated with epidemic and pandemic C:Keywords: lipid binding; lipoprotein
Query Match 15.0%; Score 689.5; DB 2; Length 1002;
Best Local Similarity: 24.2%; Pred. No. 9.3e-37;
Matches 232; Conservative 151; Mismatches 276; Indels 295; Gaps 39;
QY 4 NSAITPNTSQPINDLOGSLAAEVKFAQSQIILPAHPKEDGSDPHLTSLRKSLLLVPR-VKA 62
Db 137 NEIDYHTTEIEKKNYVSGLEGEVRFVQTHVI--SPEGRKNEPEITIGRDALILFKPSIKN 194
QY 63 DDKTPVQVEARDNNKILGTLTLYPSSLPDIYHLDGVPPEGIDFTPHNGTKIINTVA 122


```

Db      195 SSSILMKIYSEDG---LFSKVYMKSPMLPKTDQPID-----ID----- 230
QY      123 EVNKLSDAGSSIHSHLNNALVEIHTANGRWDRDLYLPQGDLECKMVRVFSAGYSST 182
Db      231 ----- 230
QY      183 VFYGDVKVTLVSGNTLLFKYVNGWFRSGELENNR-ITYAQHWSAELPAHMTVPGNLV 241
Db      231 ----- 230
QY      242 IKGNLSGRL-----NDIKIGAPGELLHTIDIGMLTTPDRDFDAKDKAHR-EYFQTI 295
Db      260 FEDEN--GNLGIIESERIKFSAPSELLIONIDGLMYKPRGNIVIKELERTAVDYFOKV 317
QY      296 PVSRLVNNYAPLHLKEVLPGLTLDMDPGNGWHSSTHQRIGKLVSHGIDNANYG 355
Db      318 PYSKILFSDYTPHFEXITLPNGTYTEKSDIGWGHQDMREAVKALVSTGINNANLG 377
QY      356 LNSTAGLGENSHPYVVAQLAAHNSRGNTANGIOVHGSGGGGIVTLDSTLGNESHEVGH 415
Db      378 IVASSGYSQ-QYNRLTNHTAHTNIGYNNGVVYHGGSGGGIVTLENTLHNSHELGH 436
QY      416 NYGLGHYVDGFKSVHRSANNNSTWGDGDKKRPINFPYSQNEKSCL--NNQCOEPF 473
Db      437 NYGLGHYVAG--GTSH---GPDTSWGDGYYKRFIANFDMKRSQSNIRPDNQEVYKPF 490
QY      474 -DGHKFGFDAMAGGSPF--SAANRFTMTNPSAIIQRFENKAVFDSRSTGFSKNAD 530
Db      491 MDKYYTLWDMAGSGYDHONGIISRYTLHHPYVARIQDMLKNGAVV---INNDFYWD-E 546
QY      531 TOEMPEYHTIDRAEQITASVNELSESKMAELAEYAVVYVHMNGNWTNRIIYPTASAD 590
Db      547 LKNIVYVYKGT-----NFKVPKIKGVVPVVTILGVYDPKIN-----PSQLYPPTY- 591
QY      591 NRGSLITIN-----HEAGYN-----SYLFING-- 612
Db      592 NYGNIFDLEKPRSESSLKGMQYKVDVNYLDRVNTHTMLVNRKEEKICRFSYLSPKGKK 651
QY      613 -----DEKVVSGYKKSFSVSDGQFWKRDVDTREARKPEQFGVPVTVLVG---Y 659
Db      652 FEFLGYEDIENKJCTGGRSIHYLEDG---KKNPI-----ESKNDYF---LSDIDGGEIS 701
QY      660 YDPEGTL-----SSYTYPAMYA-----YGFYSDSDSONLS--- 690
Db      702 YVPDSTIGESKICSLKMSGTYYGAGFIKGNSCRQIDGVFMNGFQWAFVTLNQSNGVNSTYTW 761
QY      691 DNDCOLQVDTKEGOLR-FELANHRANNTVMKPHINVPTEOPTQATLVCNNKILDTKSL 749
Db      762 SNECVLKIKDKDNNIESISIPNTRIEKNQSNKIHLSREKPIIDINVCYGEHELTSIKV 821
QY      750 TPAPEGLTYTVNGQALPAKRENEGCVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSG 809
Db      822 SDNPD-----IKLLKGPPII-----VQGEHGYT-----SYEPKLPFG 852
QY      810 AKAKVLLSDWNLN-YNRIGEFVGVNVPADMKVKVANGQYL---DFS-KPRSHRVV 861
Db      853 -----WFKHYDNFEPKPEINHELKGRVND-----NDEYICRFNFSDSOREMKFV 897

RESULT 4
B82276
ToxR-activated gene A protein VC0820 [imported] - Vibrio cholerae (strain N16961 serogro
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82276
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82276

```

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1013 <HEI>
A:Cross-references: GB:AE004167; GB:AE003852; NID:g9555268; PIDN:AAF93983.1; GSPDB:G
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0820
A:Map position: 1

Query Match      15.08; Score 689.5; DB 2; Length 1013;
Best Local Similarity 24.24; Pred. No. 9.4e-37;
Matches 232; Conservative 151; Mismatches 279; Indels 295; Gaps 39;

QY      4 NSAIYNTSPINDLOGSLAAEVKFAQSOTLPAHPKEGDSOPHLTSLRKSLLLYRP-VFA 62
Db      148 NEIDYHTTIERKNGVEGEVRFVQTHVI---SPGRKNEPEITGRALILFKPSIKN 205
QY      63 DDKTPVQVEARDNNKILGTLTPPSSLPDITVHLDGVPEGGIDFTPHNGTKKIINTVA 122
Db      206 SSSILMKIYSEDG---LTSKVYMKSPMLPKTDQPID-----ID----- 241
QY      123 EVNKLSDAGSSIHSHLNNALVEIHTANGRWDRDLYLPQGDLECKMVRVFSAGYSST 182
Db      242 ----- 241
QY      183 VFYGDVKVTLVSGNTLLFKYVNGWFRSGELENNR-ITYAQHWSAELPAHMTVPGNLV 241
Db      242 -----ENKVVSYNSYSAEIPNKKMSKNSLH 270
QY      242 IKGNLSGRL-----NDIKIGAPGELLHTIDIGMLTTPDRDFDAKDKAHR-EYFQTI 295
Db      271 FEDEN--GNLGIIESERIKFSAPSELLIONIDGLMYKPRGNIVIKELERTAVDYFOKV 328
QY      296 PVSRLVNNYAPLHLKEVLPGLTLDMDPGNGWHSSTHQRIGKLVSHGIDNANYG 355
Db      329 PYSKILFSDYTPHFEXITLPNGTYTEKSDIGWGHQDMREAVKALVSTGINNANLG 388
QY      356 LNSTAGLGENSHPYVVAQLAAHNSRGNTANGIOVHGSGGGGIVTLDSTLGNESHEVGH 415
Db      389 IVASSGYSQ-QYNRLTNHTAHTNIGYNNGVVYHGGSGGGIVTLENTLHNSHELGH 447
QY      416 NYGLGHYVDGFKSVHRSANNNSTWGDGDKKRPINFPYSQNEKSCL--NNQCOEPF 473
Db      448 NYGLGHYVAG--GTSH---GPDTSWGDGYYKRFIANFDMKRSQSNIRPDNQEVYKPF 501
QY      474 -DGHKFGFDAMAGGSPF--SAANRFTMTNPSAIIQRFENKAVFDSRSTGFSKNAD 530
Db      502 MDKYYTLWDMAGSGYDHONGIISRYTLHHPYVARIQDMLKNGAVV---INNDFYWD-E 557
QY      531 TOEMPEYHTIDRAEQITASVNELSESKMAELAEYAVVYVHMNGNWTNRIIYPTASAD 590
Db      558 LKNIVYVYKGT-----NFKVPKIKGVVPVVTILGVYDPKIN-----PSQLYPPTY- 602
QY      591 NRGSLITIN-----HEAGYN-----SYLFING-- 612
Db      603 NYGNIFDLEKPRSESSLKGMQYKVDVNYLDRVNTHTMLVNRKEEKICRFSYLSPKGKK 662
QY      613 -----DEKVVSGYKKSFSVSDGQFWKRDVDTREARKPEQFGVPVTVLVG---Y 659
Db      663 FEFLGYEDIENKJCTGGRSIHYLEDG---KKNPI-----ESKNDYF---LSDIDGGEIS 712
QY      660 YDPEGTL-----SSYTYPAMYA-----YGFYSDSDSONLS--- 690
Db      713 YVPDSTIGESKICSLKMSGTYYGAGFIKGNSCRQIDGVFMNGFQWAFVTLNQSNGVNSTYTW 772
QY      691 DNDCOLQVDTKEGOLR-FELANHRANNTVMKPHINVPTEOPTQATLVCNNKILDTKSL 749
Db      773 SNECVLKIKDKDNNIESISIPNTRIEKNQSNKIHLSREKPIIDINVCYGEHELTSIKV 832
QY      750 TPAPEGLTYTVNGQALPAKRENEGCVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSG 809
Db      833 SDNPD-----IKLLKGPPII-----VQGEHGYT-----SYEPKLPFG 863

```


A:Cross-references: GB:AE001437; PIDN:AAK80209.1; PID:gl15025253; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2252

```
Query Match      3.6%; Score 166.5; DB 2; Length 1217;
Best Local Similarity 18.6%; Pred. No. 0.018;
Matches 179; Conservative 121; Mismatches 319; Indels 343; Gaps 48;

Qy 56 LVRPKA-----DKTPQVQVARDNNKILGTLFLYPPSLP-----DTIYHLDGVPEGI 106
   ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29 IIRPVSAKAPNKTISKIKTKVENARV-----SNLSAKLNGDTLIQVINGLDETDI 79

Qy 107 -----DFPHNCKKLIINTVAEVNKLSDAGSSIHSLTNALV-----EIHFA 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 KICEPQVLKVDYKRP-SGQSSDGLVDPNKIWN-TGNISSDLSNPMTITTKMTIKIS 137

Qy 151 NGRVVRDIYLPQGDLEGKMW--RFVSSAGYSTVFYGRKVTLSVGNLTLLFKYVNGOW 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 KSDLTMSVY-----DSGKQIVKQOSTASKSVSFTHNSGDR-----FYGINGYN 181

Qy 208 FR-----SGELEN-NRITYAHIWSAELPAHIVPGLMLVIKQNLGRLND----- 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 FKEDSKGMLRNGTESYAGYQCHGSPFYNSDGYGLLVDSGSGFTIGDTSLOYSGIS 241

Qy 254 -----IKGAPGELLHTIDI-----GMLTTP-----RORFDFAKDKAARE 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 KTDYDYLMLGSPKEVISESDVSGKAPFPKATGFTNTQWNNNSLSGTGNDKDKLS 301

Qy 291 YFOT-----IPVSRMIV-----NNYAPLHLKEVMLP---TGELLTMDPGNGGWH 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 VLNTYRSKQIPIDNFCFLDFENKKGWODNYGEFKWNTDNFPDAONGQLKAYMD-SKGLKMT 360

Qy 334 GTMRQRI-----CKELVSHG-----IDNANYGLNSTAGLGENSHPYV 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 GIMPRILADSEQARYVTSKWMLPGDSASDYCSGKMMHNVFAISDV----- 409

Qy 371 VAQLAAHNSRNYANGIOVHGSGGGIVTLDTLSGNEFSHEVGHNTYGLGHY----- 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 -----RKWNWNIQ-----DAFDKGIVGFWDNDCDENYVFCNGMNNHRA 450

Qy 423 -VDGFKGSHRSAENNSTWG-----WDGDKRFPINFPYPSQTNK--- 462
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 IYDG-----QRAYKNQRWNSLRNYAGAQRYCYGMWSD-----ISTGDSMANORERM 500

Qy 463 -SCLNNOCEPFDGKFGD--AMAGSP-----PSAARNFTMTYTPNSSAIIQRFENKAV 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 LSAVN-----LGEAKWMDTGGFNDGDPPTENYARWMEFSAFTP-----IFRVHGQDNRY 550

Qy 516 -----FDSRSSTGFSKWNADQEWEPYEHTIDRAEQITA-----SVNELSEK 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 RYPWAFGSTAAAKAKAMQLRYTLIPYIYSYDRSASOSGLVRLSLMMEYPNDSNAANDK 610

Qy 559 MAELMAEYAVKVMNMGWNTNRNIVP-----TASADNRGSILTINHEAGYNSY--- 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 EAMFGDIALSPVNVQQTSEKSYLPGCNWIDITTGREYTGQ-QTINYAVDSTNNSDIP 669

Qy 608 LFING-----DEKVSOGYKKSFPVD-----GQFWKER 635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 670 LFIKSGAIPTODFENYVGERKIDTVYDAFPDOKATTFDYDDGTSYDENGSYFDQK 729

Qy 636 DVVDTRARKPQFGVPVTLVGYDYDEGTLSYI-----YP 672
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 730 MTLOTSTDSKSVQFNIDKNT--GSTPD--LRDYIVKMHVKGNGAVTANGALTYQSYSD 785

Qy 673 AMYAGYFTYSDSQNSQLDQOLQVDTKQGLRFLRANRANNTVMNKFHINTYPTESQP 732
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 786 ALKSASGEYASCT-----DIYGNVYIKVSSGDAKN-----INTVSCNPLP 826

Qy 733 TOATLVCNKK-----ILDTKSLTPAPEGLT--YTVNGOALPAKNEGCIIVSNKRYCL 785
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 827 VTITAAANPKGGTYGTPOTVSLTASKSDATIIYTL-----DGTPTVNSTK-YIA 875
```

```
Qy 786 PV 877
   : :
Db 876 PI 877

RESULT 8
JC6027
115K outer membrane protein precursor - Bacteroides thetaiotaomicron
N:Alternate names: Susc protein
C:Species: Bacteroides thetaiotaomicron
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6027
R:Reeves, A.R.; D'Elia, J.N.; Frias, J.; Salyers, A.A.
J Bacteriol. 178, 823-830, 1996
A:Title: A Bacteroides thetaiotaomicron outer membrane protein that is essential for
A:Reference number: JC6027; MUID:96146534; PMID:8550519
A:Accession: JC6027
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1038 <REE>
A:Cross-references: GB:L49338; MID:gl100064; PIDN:AAA95938.1; PID:gl100065
A:Note: It is uncertain whether Met-1, Met-14 or Met-20 is the initiator
C:Comment: This protein is an integral outer membrane protein, and it is essential f
C:Genetics:
A:Gene: susc
C:Keywords: transmembrane protein
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-1038/Product: 115K outer membrane protein #status predicted <MAT>
F:1028-1038/Domain: transmembrane #status predicted <TM>

Query Match      3.5%; Score 161; DB 2; Length 1038;
Best Local Similarity 18.1%; Pred. No. 0.031;
Matches 180; Conservative 120; Mismatches 328; Indels 366; Gaps 46;

Qy 16 NDLOGSLAAEYKFAQSQILPAEPKGDSDPHLTSLRKSLLLVR-----PVKADDKTP---V 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 NDMTGSVMA-----INPDELSKG-----ITNAQMLSGKIAGVSIINDGTPGGGA 177

Qy 69 QVEARDDNNKILGTLTPSPSLPTDIYHLDGVPGEIGDTPHNGTKRKIINTVAEVNKK-- 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 QIRIRGGSSL-----NASNDPLIVDGL---AID---NEGIKOMANGLSMVNPAD 221

Qy 127 -----LSDASGSSHTSHLTNNALVEIHTANGRWVRDIYLPQGDLEGKMTYRFVSSAGYS 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 IETLVLDASATAYIGSRASNGVITITTKGK-----NGQAPSYTYNGSVSFS 270

Qy 181 STVFYGRKVTLSVGNLTLLFKYVNGOWFPRSGELENRITYAHIWSAELPA-----HW 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 KT-----QKRYDVLSGDEYRA-----YANQLWGLKLPADLGTANTDW 307

Qy 234 -----IVPGLNLVIKOG--NLSGRNLNDIKIGAPGELLHTIDIGMLTTPRDRDFAK 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 QDQIFRTAYSTDHHSVINGGFKNLPYRS-----LGYTDDNGIVKTSNFR-RFTA 356

Qy 284 DREAUREYFQ---TIPVSRMIV---NNYAPLHLKE---VMLPTGELLTMDP---GNGG 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 SYNLAISFFEDHLKKNINAKFMNGKNRYADSCRWTRALAIDTPRPVYSNEDPTQFTGGY 416

Qy 331 WHSGTMRQRIGKELVSHGIDNANYGLNSTAGLGE-----NSHPYVVAQLAAHNSRG 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 WQN-----INSTGFSNPDWKYTSNPNSPQNPLAALKLNKDKG 454

Qy 382 NYANGIOVHGSGGGIVTLDTLSGNEFSHEVGHNYGL-----CHYVDGFKG 428
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 N-----SNDFVGNVDVYKPHFLPDLRLKHAISIGGEAEGTQT 491

Qy 429 SVHSAENNNSTGWGDG---KKRFIPNFPYSPQNEKSLNNOCEPFDGKFGFDAMA 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 492 TVSPYSFQNNYTGNGDVTQYKYNLSYIYQYI--KSLGAND-----FDLMV 538

Qy 485 GSPFSAANRFTMTYTPNSSAIIQRFENKAVPDRSSSTGFSK---WNADQEWEPYEHTI 541
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 539 GGEE-----QHFRN-----GFEEGGQGWSDYTOE---PHDA 566
```

QY 542 DRAEQITASVNELSEKMA---ELMAEYAVVYVHMGNWTRNIYIPTASADNRGSILT 597
DB 567 KUREOTAYATRNLTYSYFGRNLVSLNRYLFTTRWDGS-----SRFSKDNRW----- 615
QY 598 INHEAGYNSYLFINGDEKVVSOYKKSFFVSDQFQWKRDRVDVTDREARKPEQFGVPVTLV 657
DB 616 -----GTFPSLALGWK---IKENFLKDVNLSDKLK---LWGCITQOQ 654
QY 658 GYVDEGTLSSYIYPMYGAQF--TYSDDS--QNLSDNDQQLQVDTKEGQIRLERLANHR 713
DB 655 NIGDOPAYLPLVYVNNEXYAQYFPGDTYTSRPAFENENLKWETTTWNAGLDFGLNGR 714
QY 714 ANNTVMNKHCHINPTESQATLVCNKKILDTKSLTPAPEGLTY-----TVNGQ 763
DB 715 ITGGIDYF-----RKTWCYALRSPMNLQCPDDEYRFTGKLYGFSINAK 763
QY 764 ALPAKE-----NEGCTVSNSG--KRYCLPVG----- 788
DB 764 PIVTKDFTWDLSYNTWNNHETIKLTGGDDSDYVEAGDKISRGNTKVKQAKVGYAANS 823
QY 789 -----QBSGYSLPOMIYGOEYVYVSDGAKAKVLLSDNDLNSYNRIGEEVGNV 835
DB 824 FTVSGNNTKQVAKHVGTAANSFYIQYQYDENGKPI-----ENMFVDRNGN--GTID 874
QY 836 PADMKKYKAWNG-----QYLDFSKPRSMR 859
DB 875 SGDKVYIKKPGADVLMLGLTSKMKQYKNFDFSESLR 908
RESULT 9
S70172
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 15-Oct-1999
C:Accession: S70172; S44271
R: von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, E.; Sartingen, S.
Mol. Microbiol. 17, 313-321, 1995
A:Title: Cloning in on the toxic domain through analysis of a variant Clostridium difficile
A:Reference number: S70172; MUID:96079281; PMID:7494480
A:Accession: S70172
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2367 <VON>
A:Cross-references: EMBL:223277; NID:g761713; PIDN:CAA80815.1; PID:g761714
A:Experimental source: isolate 1470
R: Sartingen, S.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.
A:Reference number: S44271
A:Accession: S44271
A:Molecule type: DNA
A:Residues: 1-1323, 'N', 1325-2367 <SAR>
A:Cross-references: EMBL:223277
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin

Query Match 3.5%; Score 159.5; DB 2: Length 2367;
Best Local Similarity 18.4%; Pred. No. 0.14;
Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;

QY 4 NSAIYNTSQPINLQGLSAAEVKPAQSOILPAHPKFGDSQPHLSLRKSLLLVRPVAD 63
DB 1454 NSELQKNIPSYVDSEK-----ENGFINGSTKEG----LFSVSELPDVVLISKVYMD 1501
QY 64 DKTP-----VQVEARDNNKILG-----TLTLYPPSSLPTIYHLQGV 101
DB 1502 DSKPSFGYVNNLKDVKVITKDNVLTGYLKKDDIKLSLTLQDEKIKLNSVHLD-- 1559
QY 102 PEGGIDFTPHNGTKKIINTVAEYKNLSDASGSIHSLTNALVEIHTANGRWVRDIYLP 161
DB 1560 -ESG-----VAEILKPMNRKGS---TNTSDSLMSFLESNN---IKSIFV- 1596

QY 162 QGPDLGKMWRFYSSAGY---SSTVFYG-----DRKVTLSVGNT 197
DB 1597 ---NFQSNIKFILDANFIISGTTSIGQEFFICDENNNIOPIFKFTLTNTLYTVGNR 1653
QY 198 L-LFKYVNGQWFRSGELENNRIYAO-HWSAELPAHWIYVGLNLVVKQCNLSG----- 249
DB 1654 QNMIVENYLDSDGDISSTVINFSQKLYGIDSCVKNVVISNITYTDEINITYETNN 1713
QY 250 -----RLNDIKI-----CAPCELLHTIDIGMLTTPDRDP--DFAK 283
DB 1714 TYPEVIVLDANYINEKINVINLDIRYMSNDGDFILMSTSEENKVSQVIRFVNFVK 1773
QY 284 DKE-AHREYF-----QTIPVSRMIVNRYAPLHLKEVAL----- 315
DB 1774 DKTLANKLSNFSDKQDVPVSEIL-SFTPSYEDGLIGTDLGLVLYNEKFIYNNFGHM 1832
QY 316 -----PTGELLTD-----MDPGNGGHSHTMRQIRIGELVSHGI 349
DB 1833 VSGLIYINDSLYFKPPVNNLITGFTVGDGKYFNPINGGAAS-----IGETI---I 1882
QY 350 DNANYGLNSTAGLGENSHPVVVAQLAAHNSRGNY-----ANGIQVHGSGGGGIVTLDST 404
DB 1883 DDKNYFNQSG-----VLQGVFSTEDGFKYFAPAN-----TIDEN 1918
QY 405 LGNE---FSHEVGHNYGLCHYVDGFKGSVHRSANNSTMGW---DQDKRRFIPNFPY 459
DB 1919 LEGEAIDFTKLIIDENIYFEDNYRGAVE-----WKELDGMHYFSPE----- 1962
QY 459 TNEKSLNNOQEPFDGKHFGFDAMAGGSPFSAANRF--TWYTPNSAIIQRF--EN 512
DB 1963 -----TGKAFKGLNQIGDDKYYFNSDGVNQKGFVSIND 1996
QY 513 KAVFDSRS--STGFSK-----WNADTOEMEPYEHTIDRAEOITASVNELSEKMAELMA 564
DB 1997 KHYPDSDGVKMGVYTEIDGKHFFAENGQIGVFNTEGFKYFAHNEDLGNSEGEIS 2056
QY 565 EYAVVYVHMGNWTRNIYIPTASADNRGSILTNHEAGYNSYLFINGDEKVVSOYKKS 624
DB 2057 YSGIL-----NFNKKIYY---FDDSTAVVGKLEDDGSKYF---DEDTAEAVIGLS 2103
QY 625 FVSDGFQWKRDRVVDTRARKPEQFG-VPVTVLVGYVDPEGLTSSYIYPAMYGAFTYS 683
DB 2104 LINDQYIFNDDGI-----MOVGFVTINDKVFYFSDSGIIES-----GVQNI 2145
QY 684 DSDNLSDNDQQLQV---DTRKQGLRFLANHRANNTVMNKHINVPTEQATILVCN 740
DB 2146 DNYFYIDDNGIYQIGYFSDGKYFAPA-----NTVNDNIY----- 2183
QY 741 NKILDTKSLTPAPE-----GLTYVNGQALPAKENEGCIYVNSGKRYCLPVGQRS--GY 793
DB 2184 GQAVEYSGLVVRGVEDVYFGETYTIETGIWYDMENE-----SDRYFYFETKKACKGI 2236
QY 794 SLPDWIVGQEVYVSDGAKAVLLSDMDNLSY--NRIGEF-VGNVNPAD 838
DB 2237 NLIDDI---KYFDEKGMRTGLISFNNNYFNENGEIOFGYINIED 2281

RESULT 10
F97255
fusion of alpha-glucosidase (family 31 glycosyl hydrolase) and glycosidase (Trea/Mals
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97255
R: Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1157 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80833.1; PID:g15025938; GSPDB:GN00168

A: Experimental source: Clostridium acetobutylicum ATCC824

C: Genetics:

A: Gene: CAC2891

```
Query Match          3.4%; Score 156.5; DB 2: Length 1157;
Best Local Similarity 19.3%; Pred. No. 0.073;
Matches 198; Conservative 120; Mismatches 335; Indels 371; Gaps 56;

QY 60 VRADKTPQVEARDNNK-----ILGTLTPSPSLPOTIYHLDGVEGGIDFTPHNGTK 115
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 35 VRADTKVKNKSSQKQKFAKLGNGTKIKKQKDETFIICEPQVFKVDYKP-NKGS 93
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 KIINTVAEYNKLSADGSGSIHSHLTNNALVEIHTANGRWVRDI-----YLPQGPDI 166
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 SKDTLVVDNK-KWSTGNIVSSDIKSDPMV---IITKMWLKNKEDLSILVY-----DL 144
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 EGMWRFVSSGYSSTVF---YGRKRVLSVGNLTLLFYKYNQWFR---SGELEN-NRI 218
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 145 QGKLLLKQDSTASKTASFTHNSGDR-----FYGINGYFQEDSSKGLMRNGTES 193
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 TYAQHWSAELPAHWIPLVGNLVIKGNLSGRIND-----IKIGAPGEL 262
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 VTAGYGHGSGSPVWNSDGYGLVDSGGSFTIGTSLKYDGISKTDITDYVWVNGPKBI 253
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 263 LLHTIDIG-----MLTTPDRF-----DPAKDEKAHREY-FOTIPVSRMI 301
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 LSESDVSGKAPMPFKWANGFTNTQMGWDSLSGTGNDKALKSVINTYRSKQLPIDNFC 313
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 302 V-----NNYAPLHLEVLP---TGELTMDMPGNGHSGTMRORI-----GKE 343
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 LDFDWKKGQDNYGEFKWNTDNFDPSONGLKAYMD-SKGLKMTGIMKPRILADSKQGRY 372
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 LVSHG-----IDNAYGLNSTAGLGENSHYPVVAQLAAHNSRNGYANG 386
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 VTSKGNLPCDSEASDYCSKMMENVPAL-----POV-----RKWNNN 412
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 387 IQVHGSGGGGIVTLDTLGNFEFSHEVGHNYGLGHY-----VDGFKGSVHRSAMN 437
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 IQ---GAFDKGIVGF-----WNDECDENVNFGNFMNERAIYD-----QRRHN 456
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 438 NSTWG-----WDGDKKRFIPNFPSPQTNEK-----SCLNNQCEPFDGKH 477
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 QRYVSLNRNYAGAQRYSYGWMNSGD-----ISTGFDMANORERMUSAVN-----LGEAK 506
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 478 FGFD--AMAGGSP-----FSAANRFTWYTPNSSAIIORFF--ENKA-----VFDSRSSTGFS 525
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 WGHDTGFGNGDTPPENYARWMEFSAFTP-----IFRVHQDNKVYPPWAFGSTAEATK 561
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 KWNADTQEMEPYBHTIDRAEQITA-----SVNELSFESKMAELMAEYAVYKVHM 573
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 562 KAMQLRYTLIPYISYDRSASQSLGLVRSLSMMEYPNDSNAANDKEAWMFGDYMVSPV 621
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 574 WNGNWTNRNIVP-----TASADNRGSLTINHEAGYNSY---LFING----- 612
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 QEGTQSKSIVLPGNWIDYTTGREYTG--QTINIAVDSTNMSDIPFIKSGAIIPTQDFE 680
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 613 ---DEKVVSGYKKSF-----VSDGQFKERDVPVDTREARKPEQFG 650
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 681 NYVGEKIITDVYDAFPGNEASSFDYDDGTSYNTENGSYFQDKMTLERAKDLKSVQFN 740
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 651 VPVTTLVGYDPPGTLSSYIYP-----AMYGAYGTYSDDSQNLSDNOC 694
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 741 ISPKT--GYT--KSDLKNTYVRMHVKSQSDVTYVGGRRITRYASY-----DELKNAQGEY 791
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 695 QLVQDITKEGOLRRL--ANHRANNTVMKPHINVPTESQPTQATLVCKNNKIL---DTKSL 749
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 792 VVGTDITGVSVVYIKVSAGHDKN-----INVP-----CNQVOLRAYADVK-- 830
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 750 TPAPEGLTYTVNGQALPAKEN-----EGCIVSVNSGKRYCLPVGORSQSYSLPDWI 799
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 831 ----GQTYT-SQKYSLSKADSPNAIYYTLDGTAPTAVNSTK-----YTGP--- 870
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 800 VGEVYVDSKAKAVLLSDMDNLISYNRIGEFVGNVNPADMKKYKAWNGQYLDPSKPRSMR 859
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 871 ----ITDSSSKTLKFIVRD-----ANGNESDV-----FTEQITTY-----IK 903
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 860 VYVK 863
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 904 VHYK 907

RESULT 11
S54500
alpha,alpha-trehalase (EC 3.2.1.28) - yeast (Saccharomyces cerevisiae)
N:Alternate names: acid trehalase ATH1; protein YP9367.06; protein YPR026w
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 29-Oct-1999
C:Accession: S54500; S60459; S70514
R:Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54509
A:Accession: S54500
A:Molecule type: DNA
A:Residues: 1-1211 <BAD>
A:Cross-references: EMBL:Z49274; NID:g809585; PIDN:CAA89280.1; PID:g809591; MIPS:YPR(
A:Experimental source: strain AB972
R:Destruelle, M.; Holzer, H.; Klionsky, D.J.
Yeast 11, 1015-1025, 1995
A:Title: Isolation and characterization of a novel yeast gene, ATH1, that is required
A:Reference number: S60459; MUID:96076626; PMID:7502577
A:Accession: S60459
A:Molecule type: DNA
A:Residues: 131-1167, 'FAG', 1171 <DES>
A:Cross-references: EMBL:X8156; NID:g1061283; PIDN:CAA58961.1; PID:g1061284
A:Note: the authors translated the codon GTT for residue 18 as Leu, GAG for residue :
R:Alizadeh, P.; Klionsky, D.J.
FEBS Lett. 391, 273-278, 1996
A:Title: Purification and biochemical characterization of the ATH1 gene product, vacuol
A:Reference number: S70514; MUID:96350535; PMID:8764988
A:Accession: S70514
A:Molecule type: protein
A:Residues: 368-369, 'X', 371-377, 'X', 379, 'X', 381, 385-392, 'X', 394, 'X', 396 <ALI>
C:Genetics:
A:Gene: SGD:ATH1
A:Cross-references: SGD:S0006230; MIPS:YPR026w
A:Map position: 16R
C:Function:
A:Pathway: trehalose degradation
C:Keywords: glycoprotein; glycosidase; hydrolase; transmembrane protein; yeast vacuol
F:51-67/Domain: transmembrane #status predicted <TMM>

Query Match          3.4%; Score 154; DB 2: Length 1211;
Best Local Similarity 18.4%; Pred. No. 0.11;
Matches 143; Conservative 104; Mismatches 265; Indels 266; Gaps 35;

QY 74 DONNKILGTLTPSPSLPDIYHLDGVEGGIDFTPHNGTKKIINTVAEYNKLSADSGS 133
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 DDENMILGS-NLFSKNTYSROPTYVANGYIGSKI---PNIGFGYALDT---LNFYTDAPGA 173
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 SIHSHLTNNALVEIHTANGRWVRDIYLPQGPDLGKMKVRFVSSAGYSSTFYGDRKVTLS 193
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 -----LNNGWPLNHRFPAGAFVSDFYCLQ-PKLNSTNPELDDVGYSTVI-----SSIP 227
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 194 VGNLTLLFYVN-GWFRSGELENRIT-YAQHWSAELPAHWIPLVGNLVNLIKQNLGRL 251
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 QWNTLQFLSNDKWFNFQNTVLDVNTYSQ-----NLSMKDGIIVTTEL 271
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 252 N-DIKIGAPGELLH-----TIDIGMT--TPRD-----RFDPAKDEKAHREY 291
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 DMLNSQIHVKSEIWAHRHIHPLGVVSLISLTNHLPSDFDSDLDVNTWILDENFISHRTV 331
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 FOT-----IPVSRMTVNNYAPLHLEKVMPL--TGELTMDMPGNGGW 331
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 LHSITGTDKNNVPMIVQPDVNPVSSNCNAYSTCTVKYENSTNPNINSSEFEKDVSNII 391
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 -----HSGTMRQIRIGKELVSHGIDNANYGLNSTAGLGENSHYPVVAQLAA 376
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 392 NVLTEDOPKIIIVKYGIMTEFNKKEQ--DNTNIGL-----AKMIA 434
QY 377 HNSRGYANGIQVH-----
Db 435 LNSKNGYEKLLSSHKRAWYDLYNDAFTEIPSDSLEMTARSSLPHLLANTRDYNVSSDRG 494
QY 391 ---GGSG-----GGGIVTLDSTLGN-----FSHEVGHNYGLGHYVDGFKGSVHRSRSENN 437
Db 495 LPVGVSGLSDSYSGOMFMDADIWHEFALLPFFPNVAQN-----MNNYRNATHSOAKLN 548
QY 438 NSTGWDGDKRFRIPNFYPSPTNEKSLNNOCEPFDGCHKFGFDAMAGGS-----487
Db 549 AEKYGPCAIYPWTSGKYANCTSTGPCVDYEHINVDVANAFSIIYLNHGEGIDDEYLRY 608
QY 488 ---PF--SAARFMYTPT--NSAIIQRFPE-----NKAVFDSRSTGFSKKNAD 530
Db 609 TTWPIKNAAQFFATYVKNYNSLSGLYETYNLTPDEFANHINNGAFTNAGIKTILLKWATD 668
QY 531 TOEMEPYHTIDRAEQITASVYNELSESKMAELMAEYAVVYHMGNGWTRNIYIPTASAD 590
Db 669 -----TONHGEVVDPAWSEI-----SKDIYIPRSSN 696
QY 591 NRGSILTTINHEAGVNSYL-----PINGDEKVSQGYKKSFSVDSQGFWR 635
Db 697 -----ITLEY--SGMNSVEIKOADVTLMVYPLGVIN--DESILNNAIK-----DLYYYSER 744
QY 636 DVDVTPREARKPEQ--PGVPTTLGVYDPEGLTSSYIYVPMYGAIFYSDSDSONLSDN 692
Db 745 -----QSASGPANTYPPVFAAAGLLNHGSSSQSLYKSVLPYLRAPFAQFSEQSDDN 797

RESULT 12
S10317
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S10317; S21894; S22434
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.
Nucleic Acids Res. 18, 4004, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.
A:Reference number: S10317; MUID:90326540; PMID:2374729
A:Accession: S10317
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-2366 <BAR>
A:Cross-references: EMBL:X53138; NID:q40442; PIDN:CAA37298.1; PID:q40443
R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
submitted to the EMBL Data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21894
A:Molecule type: DNA
A:Residues: 1271-2366 <EIC>
A:Cross-references: EMBL:X60984; NID:q40445; PIDN:CAA43299.1; PID:q40446
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22434
A:Molecule type: DNA
A:Residues: 1791-2366 <VON>
A:Cross-references: EMBL:X60984
C:Genetics:
A:Gene: toxB
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin

Query Match 3.3%; Score 153.5; DB 2; Length 2366;
Best Local Similarity 18.6%; Pred. No. 0.35;
Matches 186; Conservative 155; Mismatches 321; Indels 339; Gaps 51;
QY 4 NSAIYPTNSQIPNDLOGSLAAEVKFAQSQILPAHPKEGDSQPHLTSRLKSLLLVPRVKAD 63

Db 1453 NSELAKNIPYFVDSSEK-----ENGFINSGTKEG-----LFVSELPDVLISKVYMD 1500
QY 64 OKTP-----VQVEARDNNKILG-----TLTYPPSSLPDTIYHLD-- 99
Db 1501 DSKPSFGYSSNNLKDVVKYTKDNVNILTYGLKDDIKISLTLQDEKTIKNSVHDES 1560
QY 100 GYPEGGIDFTPHNGTKIINTVAEVNKLSDAGSGSIHSLTNALYEHTANGRVRD-- 157
Db 1561 GVAB-----ILKPMNRKGNIN--TSDSLMSFLESNNIKSFVNFQSNIKFILDAN 1609
QY 158 -----IYLPQCPDLEGKRVFVSSAGYSTVFYDGRKVTL-----S 193
Db 1610 FIISGTTISIGFEFICDENDNIQYPIFK--NLENNYTLVGNRQNMIVEPYDLDDSGD 1668
QY 194 VNTLL-----FKYVQWFRSGELEN--NRITYAQHNSAEL---PAH-----WIVPG 237
Db 1669 ISSTVINFQSKLYG-----IDSCVKNVISPNIYTDINITPVYETNTTYPIVIVLD 1721
QY 238 LNLVIKOGNLSRLNDIKI-----GAPGELLHIDIGMLTTPDRDF--DFAKDE--AHR 289
Db 1722 ANYINEKINVN--INDLSIRYVWSNDGNDFILMSTSEENKVSQVKIRFVNVFKDKTLANK 1779
QY 290 EYF-----QTIPIVSRMIVNNYAPLHLKEVNL-----315
Db 1780 LSFNFSDKQDVPVSEIL--SFTPSYVEDGLIGYDLGLVSLYNEKFYINNFGMVSGLIYI 1838
QY 316 -----PTGELLTD-----MDPGGWHSGTMRQIRIGKELVSHGIDNANYGL 356
Db 1839 NDSLYFKPPVNNLITGFTVGDGKYFNPINGGAAS-----IGETI-----IDDKNTYF 1888
QY 357 NSTAGLGENSEHPYVVAQALAAHNSRGNY-----ANGIQVHGGSGGGIVTLDSTLGN-- 408
Db 1889 MQSG-----VLQGTGVFSTEDGFKYFAPAN-----TLDELNLEGAID 1924
QY 409 FSHEVGHNYGLHYDGFKGSVHRSRSENNNSWGW---DGDKKRFLPNFPYSPQTEKSL 465
Db 1925 FTOKLIIDENIYFDNYGAVG-----WKELDGENHYFSPE-----1961
QY 466 NNOCQEPFDGKFKGFDAMAGGSPFSAANRF--TWYTPNSSAIIQIRFF-----ENKAVFDSR 519
Db 1962 -----TGRAFKGLNIGDYKYVFNDSGVGMQKGFVSINDKNKHVFDSDS 2002
QY 520 S--STGFSK-----WNADTQEMEPYEHYIDRAEQITASVNELSESKMAELMAEYAVVKV 571
Db 2003 GVMKVGYTEIDGKHFFYFAENGEMQGVFNTEDEGFKYFAHNEDLGNKEGEISYSIGIL-- 2060
QY 572 HMNGWTRNIYIPTASADNRGSGILTIHEAGVNSYLFINGDEKVVSGYKKSFEVSDGOF 631
Db 2061 -----HFNKKIYY--FDDSFYAVGVKDLGDSKYFF--DEDTAEAYIGLSLNDGOY 2109
QY 632 WKERDVVDTRARKPEQFG--VPVTTLVGYDPEGLTSSYIYPAMYGAYGFTYSDDSQNLS 690
Db 2110 YFNDGDI-----MQVGFVITNDKVYFVSDGIIES-----GVQNIIDNNIYFI 2151
QY 691 DNDQCOLV---DTKEQLRFLRANHANNTYMNKFINVPTESQPTQATLVCCNKILDTK 747
Db 2152 DDNGIYQIGVIGVTSDDGKYKYPAPA-----NTVNDNIY-----GOAVEYS 2189
QY 748 SLTPAPE-----GLTYTVNGQALPAKENECCIVSVNSGKRYCLPVGORS--GYSIPDWTIV 800
Db 2190 GLYRVGEDVYFGETYTITGTWIIDENE-----SDKYFNPETKKACKGINLIDDI- 2241
QY 801 GQEVYVDSGAKAVLLSDMDNLISY--NRIGEF--VGNVNPAD 838
Db 2242 --KYFDEKIGMETGLISFENNYYFENGENGEMQFGYINIED 2280

RESULT 13
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83412
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; NID:99947856; PIDN:AAG05263.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874

Query Match 3.2%; Score 148.5; DB 2; Length 2468;
Best Local Similarity 19.5%; Pred. No. 0.79;
Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 10 NTSQPINLQSLAEVFAQSILP-----AHP-----KEGDSQPH----- 46
DB 324 NNSTPVTVEADTTAPADTDQVAPDSSVTGNAEPGATVGVDTDGGQDDTTVVVGGP 383
QY 47 -----LTSRLKSLLRPVKADKTPQVEARD-----DNNKIL----- 80
DB 384 GSFEVPLNPPLNGETVTVITDPAGNSSTPTVTAEPDPAQPVNASGSLVSGTAEAG 443
QY 81 -----GTLTLYPPSLPD-----TIYHLDGV 101
DB 444 VTIVITDGNPNIGOTSADANGNSFTPCGSLPDGVVNVVVARDAAGNSSPATISITVDGV 503
QY 102 PEGGIDFTPHNGTKKIINTVAEVNK---LSASGSSIHSHLTNNALVEIHTANGRWVVDI 158
DB 504 APNAPVPSNGSE--LSGTAEPSGVTLTDGNGNPIGQTAD-----ANGNW---S 550
QY 159 YLPGQPDLEGKMYRVS--SAGYSS-----TVFYGDRKVTLSVGNWTLFFKYVNGOMF 208
DB 551 FTFPTPLPDGTVVNVVARDAGNSPPASVTVDVAPATPTVDPNSGTTL----- 600
QY 209 RSGELENRITVAQHSWSELPAHWIVPGLNLVKGSLGSLNDIKIGAPGELLHTID 268
DB 601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 269 IGMATPRDRFPAKKEAHREYFOTIPVSRVINYVAPLHLKEYMLPTGELLTD----- 323
DB 620 IQGVTA-----DGSNGWTFSTPLPNCVTNNATAD-----PSGNASSPASVT 664
QY 324 -----MDPGNGHSGTMRQRIKGLVSHGIDNANYGLNSTAGIGENSHPHYVAQL 374
DB 665 DAVAPATPVVNPNSGTTLSGT-----AEPGATVTLTDG--NGNP--IQGV 705
QY 375 AAHNSRGNYA-----NGIOVH-----GGSGGGGIVTLDS-----TL 405
DB 706 TADGS--GNWSPFTPTPLPNCVTNNATADASNTSAGSSVTVDVAPATPVINPSNGTTL 764
QY 406 GNEFSHEVGHNGYGLHYDGFSGVHRSANNNSTWGDGKRKRPINFPYS--QTNKSC 464
DB 765 SG--TAEPGSSVTL--TDGNGNPIQGVTAADGSGNW-----SFTSTPLADGT 808
QY 465 LNNQCEPFDGKHFDAAGGSPFSAANRFTWYTPNSSALIQRFFENKAVTDSRSSTGF 524
DB 809 VNATATDP-----AGNTSGOGSTTVDGVAFTPTV-----NLSNGSLSGTA-- 850
QY 525 SKWNADTQEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVVYHVMNNGNWTNRYI 584
DB 851 -----EPGS-----TVILTDCNGNPIAEVTAD-----GSGNWT---YT 880
QY 585 PTASADNRGSIILTINHEAGYNSYLFINGDEKVVYSGYKKSFYSDGQFWRKEROVDVTEAR 644
DB 881 PSTPIANGTVNVVVAODAGNS-----SPGASVT-----VDSQAPA 916
QY 645 KPGQFGVPVTVTLGVYDPEGTLSYIYPAMYGAFTYSDDSQNLSDNDQCLOQVDTKEGO 704

DB 917 APVNPSTGTTLSGTAEPCATV-----TLTDGNGNFIG---QVTAD--GSGN 958
QY 705 LRFRLANRANNTVMNKHFNHPTESOPTQATLVCCNKLDTKSLTPAPEGLTYVNGQA 764
DB 959 WSTPTCTPLANGTVVN-----ATASDPTGNTSAPASTTVD--SVAPAAP--VVPNSGAE 1009
QY 765 LPAKENEGCIVSVNSCKRYCLPVQBSGSYSLPDM 798
DB 1010 ISGTAEPCATVTLTDGSGN--PIGQVTAADGSGNW 1041

RESULI 14
I40805
collagenase - Clostridium histolyticum
C:Species: Clostridium histolyticum
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 21-Jul-2000
C:Accession: I40805; T44355
R:Yoshihara, K.; Matsushita, O.; Minami, J.; Okabe, A.
J. Bacteriol. 176, 6489-6496, 1994
A:Title: Cloning and nucleotide sequence analysis of the colH gene from Clostridium h
A:Reference number: I40805; MUID:95050206; PMID:7961400
A:Accession: I40805
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1021 <RES>
A:Cross-references: GB:D29981; NID:9563954; PIDN:BAA06251.1; PID:q710023
R:Matsushita, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.
J. Bacteriol. 181, 923-933, 1999
A:Title: Gene duplication and multiplicity of collagenases in Clostridium histolyticu
A:Reference number: 222752; MUID:99121032; PMID:9922257
A:Accession: T44355
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1021 <MAT>
A:Cross-references: EMBL:AB014075; NID:93688863; PIDN:BAA34542.1; PID:g3892646
A:Experimental source: strain JCM 1403
C:Genetics:
A:Gene: colH
C:Superfamily: microbial collagenase

Query Match 3.2%; Score 148; DB 2; Length 1021;
Best Local Similarity 18.6%; Pred. No. 0.22;
Matches 169; Conservative 121; Mismatches 330; Indels 288; Gaps 46;

QY 74 DDNNKILGTLTPSSLPDTHVLDGVPBGDDFTPHNGTKKIINTVAEVNKLSDASGS 133
DB 263 DDNNSWI-----IDNGIYHI--APLGLK-----HSNNKIGIETLTVMK----- 299
QY 134 SIHSLTNALVEIHTANGRWVVDIYLPQGPDLLEGKW---RVSAGYS---STVFYGD 187
DB 300 -VYPYLSMQ-----HLOSADQIKRHY--DSKDAEGNKIPLDKFKKGGKCYKPTTFDD 351
QY 188 RKVTLSCVNTL-----LFKYNGWQFRSGELENRITVAQHSWSELPAHWIVPG 237
DB 352 GKVIKAGARVEEKYKRLYASKEVNSQFPR-----VYG 386
QY 238 LNLVKGNGSLGRNDIKIGAPGELLHTI-----DIGMLTTPDRFRFAKKEAHR 29C
DB 387 IDKPLEGPNDDILTVIYNSPEYKLSVLYGVDYTNNGMYTEPGTF--FTYEREAQS 445
QY 291 YFQTIPTVSRMIVNYY-----APLHLKEYMLPTCELLTMDPGNGHSGTMRQRIKGL 344
DB 446 TYTLELFRHEYTHYLOGRYAVFGWGRTKLYNDRLTWTEG-----GAE 492
QY 345 VSHGIDNANYGLNSTAGLGENSHPHYVAQLAAHNSRGN--YANGIOVHSGSGGGIVTLD 402
DB 493 FA-----GSTRTSGILPRKS--IVSNI--HNTTRNNRYKLSDTVH----- 528
QY 403 STLGNFSEVGHNYGLHYVD-----GFKGSVHRSANNNSTWGDGKRKRPINFY 455
DB 529 SKYGASFEF---YNYAC--MFMQDMYKMDGILNKLNDLAKNNDV--GYDN-----Y 574

